

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:14:14 ; Search time 5 Seconds
(without alignments)
282.548 Million cell updates/sec

Title: US-10-057-890A-16
Perfect score: 28
Sequence: 1 GGGGS 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	28	100.0	5	2	AAR72707	Aar72707 Linker fo
2	28	100.0	5	2	AAR34034	Aar34034 Linking s
3	28	100.0	5	2	AAR95062	Aar95062 scFv spac
4	28	100.0	5	2	AAW17094	Aaw17094 Gly(4)-Se
5	28	100.0	5	2	AAW19543	Aaw19543 Chimeric
6	28	100.0	5	2	AAY02127	Aay02127 Peptide l
7	28	100.0	5	2	AAY25357	Aay25357 IFNAR2/IF
8	28	100.0	5	2	AAY33597	Aay33597 VH-VL dom
9	28	100.0	5	2	AAY43496	Aay43496 Linker fo

10	28	100.0	5	3	AAy83210	Aay83210	Peptide l
11	28	100.0	5	3	AAB06226	Aab06226	Expressio
12	28	100.0	5	3	AAy54917	Aay54917	Linker fr
13	28	100.0	5	3	AAy43750	Aay43750	Linker us
14	28	100.0	5	3	AAB14535	Aab14535	Peptide l
15	28	100.0	5	3	AAB00156	Aab00156	Linker us
16	28	100.0	5	4	AAE06269	Aae06269	Glycine l
17	28	100.0	5	4	AAE11873	Aae11873	Spacer pe
18	28	100.0	5	4	AAB57374	Aab57374	Flexible
19	28	100.0	5	4	ABB56480	Abb56480	Human sin
20	28	100.0	5	4	AAG62343	Aag62343	Linker pe
21	28	100.0	5	4	AAB50848	Aab50848	Fluoresce
22	28	100.0	5	4	AAB57784	Aab57784	Flexible
23	28	100.0	5	4	AAB55527	Aab55527	Peptide l
24	28	100.0	5	4	AAB81033	Aab81033	Linker pe
25	28	100.0	5	4	AAG62613	Aag62613	Linker pe
26	28	100.0	5	4	AAU05173	Aau05173	Pain-reli
27	28	100.0	5	4	AAU05172	Aau05172	Pain-reli
28	28	100.0	5	4	AAG63011	Aag63011	Peptide l
29	28	100.0	5	4	AAB68571	Aab68571	Linker pe
30	28	100.0	5	4	AAB74580	Aab74580	Context-d
31	28	100.0	5	4	AAE08730	Aae08730	Peptide l
32	28	100.0	5	5	ABG31871	Abg31871	Spacer pe
33	28	100.0	5	5	AAU76043	Aau76043	Synthetic
34	28	100.0	5	5	ABB79971	Abb79971	Linker pe
35	28	100.0	5	5	ABB79970	Abb79970	Linker pe
36	28	100.0	5	5	ABP52872	Abp52872	Flexible
37	28	100.0	5	5	ABG31035	Abg31035	Binding d
38	28	100.0	5	5	ABJ05006	Abj05006	VH and VL
39	28	100.0	5	5	ABP48193	Abp48193	Peptide l
40	28	100.0	5	5	ABG60659	Abg60659	Polyimmun
41	28	100.0	5	5	AAO14387	Aao14387	Linker pe
42	28	100.0	5	5	AAE19822	Aae19822	Linker pe
43	28	100.0	5	5	AAE25959	Aae25959	Linker pe
44	28	100.0	5	5	ABJ03945	Abj03945	Peptide l
45	28	100.0	5	5	AAU75495	Aau75495	Human Gly

ALIGNMENTS

RESULT 1

AAR72707

ID AAR72707 standard; peptide; 5 AA.

XX

AC AAR72707;

XX

DT 31-OCT-1995 (first entry)

XX

DE Linker for apo A-I and apo B-100 fusion polypeptide.

XX

KW Apo A-I; LDL cholesterol; low density lipoprotein; fusion polypeptide;
KW linker.

XX

OS Synthetic.

XX

PN US5408038-A.

XX
 PD 18-APR-1995.
 XX
 PF 08-OCT-1992; 92US-00959946.
 XX
 PR 09-OCT-1991; 91US-00774633.
 PR 18-JUN-1992; 92US-00901706.
 XX
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI Witztum JL, Koduri KR, Young SG, Smith RS, Curtiss LK;
 XX
 DR WPI; 1993-134378/16.
 XX
 PT Polypeptide mimic of native apo B-100 and native apo A-I - useful in
 PT assays for LDL and HDL in plasma samples.
 XX
 PS Disclosure; Col 18; 4lpp; English.
 XX
 CC A dispersible apo A-I/B-100 fusion polypeptide is claimed which contains
 CC a first AA sequence of apo A-I (see AAR72605) and that includes at least
 CC AA sequence positions 120-135 (see AAR72606). The two sequences are
 CC operatively linked. An exemplary linking sequence is AAR72707 whose
 CC encoding DNA can be ligated between an apo A-I and a B-100 encoding DNA
 CC sequence
 XX
 SQ Sequence 5 AA;

 Query Match 100.0%; Score 28; DB 2; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 GGGGS 5
 |||||
 Db 1 GGGGS 5

RESULT 2

AAR34034

ID AAR34034 standard; protein; 5 AA.

XX

AC AAR34034;

XX

DT 25-MAR-2003 (revised)

DT 13-AUG-1993 (first entry)

XX

DE Linking sequence whose encoding DNA can be ligated between an apo A-I-
 DE and a B-100-encoding DNA sequence.

XX

KW Lipoprotein; apoprotein; B-100; A-I; LDL; HDL; assay.

XX

OS Synthetic.

XX

PN WO9307165-A1.

XX

PD 15-APR-1993.

XX

PF 09-OCT-1992; 92WO-US008634.

XX

PR 09-OCT-1991; 91US-00774633.

PR 18-JUN-1992; 92US-00901706.

PR 08-OCT-1992; 92US-00959946.

XX

PA (SCRI) SCRIPPS RES INST.

XX

PI Smith RS, Curtiss LK, Koduri KR, Witztum JL, Young SG;

XX

DR WPI; 1993-134378/16.

XX

PT Polypeptide mimic of native apo B-100 and native apo A-I - useful in
PT assays for LDL and HDL in plasma samples.

XX

PS Disclosure; Page 14 and page 35; 137pp; English.

XX

CC The inventors claim a portion of the polypeptide contg. apo B-100 that
CC immunoreacts with antibodies secreted by the hybridoma MB47 having ATCC
CC Accession No. 8746. Polypeptides specifically claimed include residues
CC 217-297, 216-310, 216-331, 216-352, 216-377, 1-377, 205-297, 173-297, 140
CC -297. DNA sequences encoding the polypeptides are also claimed. Also
CC claimed are a fusion polypeptide that contains: (a) a first amino
CC acid residue sequence up to 250 residues in length that includes residues
CC 120-135 of apo A-I, (b) a second amino acid residue sequence up to 375
CC residues in length that includes residues 217-297 of apo B-100 and DNA
CC encoding it. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25
CC -MAR-2003 to correct PR field.)

XX

SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 2; Length 5;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGS 5

|||||

Db 1 GGGGS 5

RESULT 3

AAR95062

ID AAR95062 standard; peptide; 5 AA.

XX

AC AAR95062;

XX

DT 18-AUG-1996 (first entry)

XX

DE scFv spacer peptide.

XX

KW Nucleic acid transfer system; gene transfer; gene therapy;
KW cell targeting; multidomain protein; vector; cancer; scFv;
KW single chain antibody.

XX

OS Synthetic.

XX

PN WO9613599-A1.

XX
 PD 09-MAY-1996.
 XX
 PF 31-OCT-1995; 95WO-EP004270.
 XX
 PR 01-NOV-1994; 94EP-00810627.
 XX
 PA (WELS/) WELS W.
 XX
 PI Wels W, Fominaya J;
 XX
 DR WPI; 1996-239505/24.
 XX
 PT Nucleic acid transfer system for gene therapy, e.g. against cancer -
 PT includes toxin translocation domain to target nucleic acid to specific
 PT cell.
 XX
 PS Disclosure; Page 8; 106pp; English.
 XX
 CC A flexible spacer peptide (AAR95062) is used to link the light chain
 CC variable domain to the heavy chain variable domain of a single chain
 CC recombinant antibody (scFv). The scFv may be derived from a monoclonal
 CC antibody, e.g. MAb FRP5, and forms the ligand domain of a multidomain
 CC protein (see also AAR95053 and AAR95056-58) that is used with an effector
 CC nucleic acid in a novel nucleic acid transfer system suitable for gene
 CC therapy. The ligand domain has a target cell recognition function and
 CC allows cellular internalization of the multidomain protein/nucleic acid
 CC complex
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 2; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 1 GGGGS 5

RESULT 4

AAW17094

ID AAW17094 standard; peptide; 5 AA.

XX

AC AAW17094;

XX

DT 14-SEP-1999 (first entry)

XX

DE Gly(4)-Ser linker peptide for chimeric protein construct.

XX

KW Haematopoietic protein; human; granulocyte-colony stimulating factor;
 KW G-CSF; interleukin; c-mpl ligand; linker; gene therapy; aplastic anaemia;
 KW stem cell expansion; leukopaenia; neutropaenia; vector; bone marrow;
 KW thrombocytopaenia; blood cell activation; growth.

XX

OS Synthetic.

XX

PN WO9712985-A2.
 XX
 PD 10-APR-1997.
 XX
 PF 04-OCT-1996; 96WO-US015774.
 XX
 PR 05-OCT-1995; 95US-0004834P.
 XX
 PA (SEAR) SEARLE & CO G D.
 XX
 PI Feng Y, Staten NR, Baum CM, Summers NL, Caparon MH, Bauer SC;
 PI Zurfluh L, Mckearn JP, Klein BK, Lee SC, Mcwherter CA, Giri JG;
 XX
 DR WPI; 1997-226228/20.
 XX
 PT Multi-functional haematopoietic receptor agonists - used to stimulate the
 PT production of haematopoietic cells in patients.
 XX
 PS Disclosure; Page 33; 616pp; English.
 XX
 CC The invention relates to a novel haematopoietic protein (HP) comprising
 CC an amino acid (AA) sequence of formula: R1-L1-R2; R2-L1-R1; R1-R2; or R2-
 CC R1; where R1 and R2 are independently selected from: (I) a modified human
 CC granulocyte-colony stimulating factor (hG-CSF) AA sequence; (II) a
 CC modified human interleukin-3 (hIL-3) AA sequence; (III) a modified human
 CC c-mpl ligand; and a colony stimulating factor (CSF); and L1 = a linker
 CC capable of linking R1 to R2. This sequence represents an example of a
 CC linker used to construct the proteins of the invention. Vectors
 CC comprising the nucleic acid molecules are useful for the recombinant
 CC production of HP. The nucleic acid molecules are useful in gene therapy.
 CC The HP's are useful for stimulating the production of haematopoietic
 CC cells in patients, selective ex vivo expansion of stem cells and for
 CC treatment of haematopoietic disorders. Disorders that can be treated
 CC include leukopaenia, neutropaenia, aplastic anaemia and
 CC thrombocytopaenia. In vitro uses include the ability to stimulate bone
 CC marrow and blood cell activation and growth before infusion into the
 CC patients
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 2; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 1 GGGGS 5

RESULT 5
 AAW19543
 ID AAW19543 standard; peptide; 5 AA.
 XX
 AC AAW19543;
 XX
 DT 19-FEB-1998 (first entry)
 XX

DE Chimeric protein pentapeptide linker for the MBP moiety and PE moiety.
 XX
 KW Pseudomonas exotoxin; myelin basic protein; chimeric protein;
 KW autoimmune disease; multiple sclerosis; human.
 XX
 OS Synthetic.
 XX
 PN WO9719179-A1.
 XX
 PD 29-MAY-1997.
 XX
 PF 17-NOV-1996; 96WO-IL000151.
 XX
 PR 17-NOV-1995; 95IL-00116044.
 PR 26-DEC-1995; 95IL-00116559.
 XX
 PA (YISS) YISSUM RES & DEV CO.
 XX
 PI Lorberboum-Galski H, Steinberger I, Beraud E, Marianovsky I;
 PI Yarkoni S;
 XX
 DR WPI; 1997-298116/27.
 XX
 PT New Pseudomonas exotoxin-myelin basic protein chimeric proteins - used
 PT for the treatment of auto:immune diseases, particularly multiple
 PT sclerosis.
 XX
 PS Claim 6; Page 29; 54pp; English.
 XX
 CC New chimeric proteins have been developed comprising a Pseudomonas
 CC aeruginosa exotoxin (PE) moiety linked to a myelin basic protein (MBP)
 CC moiety selected from: (a) MBP; (b) amino acids 69-88 of guinea-pig MBP or
 CC an antigenic portion; (c) amino acids 84-102 of human MBP or an antigenic
 CC portion; (d) amino acids 143-168 of human MBP or an antigenic portion;
 CC and (e) an amino acid sequence in which one or more amino acids have been
 CC deleted, added, substituted or mutated in the amino acid sequences of
 CC (a), (b), (c), or (d), the modified sequences retaining at least 75%
 CC homology with the amino acid sequences. The present sequence represents
 CC the preferred pentapeptide linker used to link the MBP moiety and PE
 CC moiety in a chimeric protein. The chimeric proteins can be used for the
 CC treatment of autoimmune diseases such as multiple sclerosis. The chimeric
 CC proteins can specifically target and kill MBP specific T cells while
 CC having no effect on non-target cells
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 2; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 1 GGGGS 5

RESULT 6
 AAY02127

ID AAY02127 standard; protein; 5 AA.
 XX
 AC AAY02127;
 XX
 DT 16-JUL-1999 (first entry)
 XX
 DE Peptide linker used to make multifunctional proteins.
 XX
 KW Angiostatin; endostatin; interferon; thrombospondin;
 KW interferon-inducible protein; platelet factor 4; anti-angiogenic;
 KW anti-tumor; multifunctional protein; angiogenic-mediated disease; cancer;
 KW diabetic retinopathy; macular degeneration; arthritis;
 KW tumor cell production; peptide linker.
 XX
 OS Homo sapiens.
 XX
 PN WO9916889-A1.
 XX
 PD 08-APR-1999.
 XX
 PF 30-SEP-1998; 98WO-US020464.
 XX
 PR 01-OCT-1997; 97US-0060609P.
 XX
 PA (SEAR) SEARLE & CO G D.
 XX
 PI Bolanowski MA, Caparon MH, Casperson GF, Gregory SA, Klein BK;
 PI Mckearn JP;
 XX
 DR WPI; 1999-255098/21.
 XX
 PT New multifunctional proteins useful for treating angiogenic-mediated
 PT diseases.
 XX
 PS Disclosure; Page 111; 121pp; English.
 XX
 CC The specification describes multifunctional proteins which comprise
 CC combinations of angiostatin, endostatin, interferon, thrombospondin,
 CC interferon-inducible protein and platelet factor 4, and have anti-
 CC angiogenic and/or anti-tumor activity. The multifunctional protein may
 CC exhibit useful properties such as having similar or greater biological
 CC activity when compared to a single factor or by having improved half-life
 CC or decreased adverse side effects, or a combination of these properties.
 CC The proteins can be used for treating an angiogenic-mediated disease,
 CC e.g. cancer, diabetic retinopathy, macular degeneration, or arthritis.
 CC They can also be used for inhibiting the production of tumor cells
 CC (characteristic of lung, breast, ovarian, prostate, pancreatic, gastric,
 CC colon, renal, bladder cancers; melanoma, hepatoma, sarcoma and lymphoma)
 CC in a patient and for inhibiting tumor growth. AAY02125-32 represent
 CC peptide linkers used to make the multifunctional proteins of the
 CC invention
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 2; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
| | | |
Db 1 GGGGS 5

RESULT 7

AA25357

ID AAY25357 standard; peptide; 5 AA.

XX

AC AAY25357;

XX

DT 06-SEP-1999 (first entry)

XX

DE IFNAR2/IFN-beta complex peptide fragment 1.

XX

KW IFNAR2; IFN-beta; type I interferon; IFNAR/IFN complex; IFN; antiviral;
KW human interferon alpha/beta receptor; anticancer; immunomodulatory;
KW anti-arthritic; antidiabetic; treatment; hepatitis; viral infection;
KW hairy cell leukemia; Kaposi's sarcoma; multiple myeloma; cancer; lupus;
KW diabetes; multiple sclerosis; rheumatoid arthritis; myasthenia gravis;
KW acquired immune deficiency syndrome.

XX

OS Synthetic.

XX

PN WO932141-A1.

XX

PD 01-JUL-1999.

XX

PF 18-DEC-1998; 98WO-US026926.

XX

PR 19-DEC-1997; 97US-0068295P.

XX

PA (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.

PA (MCIN/) MCINNIS P G.

XX

PI Tepper M, Cunningham M, Sherris D, El Tayar N, McKenna S;

XX

DR WPI; 1999-405115/34.

XX

PT Prolonging in vivo activity of type I interferon by complexing.

XX

PS Example 8; Page 76; 86pp; English.

XX

CC This invention describes a novel method for prolonging the in vivo effect
CC of type I interferon (IFN) by administering IFN as a complex (A) with a
CC subunit (I) of the human interferon alpha/beta receptor (IFNAR). The
CC product of the invention has antiviral, anticancer, immunomodulatory,
CC anti-arthritic and antidiabetic activity. (A) have the antiviral,
CC anticancer and immunomodulating activities of IFN, e.g. for treating
CC hepatitis and other viral infections, hairy cell leukemia, Kaposi's
CC sarcoma, multiple myeloma and other cancers, multiple sclerosis,
CC rheumatoid arthritis, myasthenia gravis, diabetes, acquired immune
CC deficiency syndrome and lupus. When complexed in (A), the storage life of
CC IFN is increased (i.e. it is stabilized against oligomerization, without
CC the need for storage at acidic pH) and its biological effect is
CC potentiated

XX

SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 8

AAAY33597

ID AAY33597 standard; protein; 5 AA.

XX

AC AAY33597;

XX

DT 20-DEC-1999 (first entry)

XX

DE VH-VL domain linker peptide #9.

XX

KW Antigen binding; single chain; variable domain; VH domain; light chain;
KW heavy immunoglobulin chain; VL domain; anticancer; antiviral; tumor;
KW antibacterial; antimalarial; antiinflammatory; treatment; prevention;
KW diagnosis; vaccine; autoimmune disease; inflammation; blood disorder;
KW transplant rejection; arthritis; nervous system disorder; infection.

XX

OS Synthetic.

XX

PN DE19816141-A1.

XX

PD 14-OCT-1999.

XX

PF 09-APR-1998; 98DE-01016141.

XX

PR 09-APR-1998; 98DE-01016141.

XX

PA (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.

XX

PI Kontermann R, Sedlacek H, Mueller R;

XX

DR WPI; 1999-581511/50.

XX

PT New polyspecific binding agents containing variable heavy and light
PT constructs connected via peptide linker, used for treatment, prevention
PT or diagnosis of e.g. cancer.

XX

PS Claim 7; Page 17; 20pp; German.

XX

CC This sequence represents a novel single-chain molecule (I) that binds
CC multiple antigens and comprises two variable domains of heavy
CC immunoglobulin chains (VH), having specificities A and B and two variable
CC domains of light chains (VL), also with specificities A and B. The
CC domains are provided as two VH-VL constructs which are attached via a
CC peptide (P). Any VH and VL may be replaced by their functional fragments.
CC The products of the invention have anticancer, antiviral, antibacterial,

CC antimalarial and antiinflammatory activity. (I) are used to treat,
 CC prevent or diagnose tumors (e.g. as tumor vaccines), autoimmune diseases
 CC and inflammation (e.g. transplant rejection and arthritis), blood
 CC disorders (e.g. of the coagulation and/or circulatory systems, such as
 CC anemia, leucopenia, thrombocytopenia and hypertension), nervous system
 CC disorders and/or infections (by viruses or bacteria, or malaria),
 CC including, when (I) include a fusogenic peptide, use for gene transfer.
 CC (I) are produced simply and in predominantly homogeneous form, in a wide
 CC variety of hosts, either in secreted or membrane-bound forms. This
 CC sequence represents a VH-VL domain linker peptide which is used to
 CC illustrate the method of the invention
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 2; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 1 GGGGS 5

RESULT 9

AA43496

ID AA43496 standard; peptide; 5 AA.

XX

AC AA43496;

XX

DT 26-JAN-2000 (first entry)

XX

DE Linker for dual avb3 receptor/metastasis-associated receptor ligands.

XX

KW Interferon-alpha-2b; IFN-alpha; avb3 antagonist; avb3 receptor ligand;
 KW metastasis-associated receptor ligand; angiogenesis; cell proliferation;
 KW anti-angiogenic protein; avb3-integrin; cancer; arthritis;
 KW macular degeneration; diabetic retinopathy; hemangioma; psoriasis;
 KW osteoporosis; thrombosis; angina; atherosclerosis; antiviral;
 KW antibacterial; antifungal.

XX

OS Homo sapiens.

XX

PN W09951638-A1.

XX

PD 14-OCT-1999.

XX

PF 07-APR-1999; 99WO-US004295.

XX

PR 08-APR-1998; 98US-0081074P.

XX

PA (SEAR) SEARLE & CO G D.

XX

PI Tjoeng FS, Fok KF;

XX

DR WPI; 1999-620196/53.

XX

PT New conjugates of integrin antagonist and ligand for metastasis-

PT associated receptor, for treating angiogenesis-related diseases, e.g.
PT cancer.
XX
PS Claim 18; Page 86; 108pp; English.
XX
CC The present sequence represents a linker used to join the avb3 antagonist
CC and the metastasis-associated receptor ligand, in the pharmaceutical
CC compounds of the invention. These compounds are dual avb3
CC receptor/metastasis-associated receptor ligands, and inhibit angiogenesis
CC and thus proliferation of (cancer) cells. One component binds to the avb3
CC receptor and the other to a metastasis-associated receptor. The avb3
CC antagonists may also be conjugated to anti-angiogenic proteins, such as
CC IFN-alpha and its derivatives. The compounds are used to treat
CC angiogenesis-related disorders (mediated by the avb3-integrin),
CC specifically cancer (of lung, breast, ovary, prostate, stomach, colon,
CC kidney or bladder, also melanoma, hepatoma, sarcoma and lymphoma),
CC arthritis and macular degeneration, and also diabetic retinopathy,
CC hemangioma, psoriasis, osteoporosis, thrombosis, angina, atherosclerosis
CC etc. The compounds may also be useful as antiviral, antibacterial and
CC antifungal agents
XX
SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
| | | | |
Db 1 GGGGS 5

RESULT 10

AAAY83210

ID AAY83210 standard; peptide; 5 AA.

XX

AC AAY83210;

XX

DT 24-JUL-2000 (first entry)

XX

DE Peptide linker used in construction of a_vb_3 integrin/IFN alpha.

XX

KW Biconjugate; a_vb_3 integrin; interferon alpha; angiogenesis; cancer;

KW tumour; osteoporosis; Paget's disease; Karposi's sarcoma;

KW periodontal disease; metastasis; neoplasia; retinopathy; arthritis;

KW psoriasis; leukaemia; malignant melanoma; atherosclerosis;

KW smooth muscle cell migration; inhibition; treatment; antagonist; angina;

KW thrombosis; restenosis; antiviral; antifungal; antibacterial.

XX

OS Synthetic.

XX

PN WO200009143-A1.

XX

PD 24-FEB-2000.

XX

PF 07-APR-1999; 99WO-US004296.

XX

PR 13-AUG-1998; 98US-0096442P.

XX

PA (SEAR) SEARLE & CO G D.

XX

PI Fok KF, Tjoeng FS;

XX

DR WPI; 2000-205894/18.

XX

PT New bioconjugates comprising an avb3 antagonist and a metastatic-
PT associated receptor ligand, useful for treating cancer and other
PT angiogenic diseases, or as antiviral, antifungal or antibacterial agents.

XX

PS Claim 19; Page 88; 123pp; English.

XX

CC Bioconjugates comprising one or more a_vb_3 antagonist moieties coupled
CC to a peptide or polypeptide having anti-angiogenic properties can be used
CC for treating a human patient with an angiogenesis-mediated disease, e.g.
CC cancer, arthritis, or macular degeneration. The a_vb_3 integrin is
CC normally associated with endothelial cells but can promote the formation
CC of blood vessels (angiogenesis) in tumours. The a_vb_3 integrin is also
CC known to play a role in tumour metastasis, neoplasia, osteoporosis,
CC Paget's disease, retinopathy, arthritis, periodontal disease, psoriasis
CC and smooth muscle cell migration. Interferon alpha is a family of
CC proteins which possess complex antiviral, antineoplastic and
CC immunomodulating activities. Interferon alpha is effective against a
CC variety of cancers including hairy cell leukaemia, chronic myelogenous
CC leukaemia, malignant melanoma and Kaposi's sarcoma. Multi-functional
CC bioconjugates comprising both a_vb_3 antagonists and interferon alpha 2b
CC can exhibit greater biological activity when compared to a single factor
CC or having improved half-life or decreased adverse side effects, or a
CC combination of these properties. They can be used for inhibiting elevated
CC levels of tumor antigens, inhibiting the proliferation of tumor cells and
CC inhibiting tumor growth. The bioconjugates can also be used for treating
CC e.g. osteoporosis, humoral hypercalcemia of malignancy, Paget's disease,
CC retinopathy including diabetic retinopathy, arthritis, including
CC rheumatoid arthritis, periodontal disease, psoriasis, thrombosis, angina,
CC atherosclerosis, smooth muscle cell migration and restenosis in a mammal.
CC They are also useful as antiviral, antifungal and antibacterial agents.
CC This sequence is a peptide linker used in the construction of the multi-
CC functional bioconjugates

XX

SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 3; Length 5;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5

|||||

Db 1 GGGGS 5

RESULT 11

AAB06226

ID AAB06226 standard; peptide; 5 AA.

XX

AC AAB06226;

XX
 DT 22-NOV-2000 (first entry)
 XX
 DE Expression vector CANTAB5E inserted peptide.
 XX
 KW Modified RNase; eosinophil derived neurotoxin protein; EDN; cancer;
 KW Kaposi's sarcoma; neoplastic endothelial cell;
 KW non-neoplastic endothelial cell; expression vector.
 XX
 OS Synthetic.
 XX
 PN WO200026233-A1.
 XX
 PD 11-MAY-2000.
 XX
 PF 01-NOV-1999; 99WO-US025737.
 XX
 PR 02-NOV-1998; 98US-0106732P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Rybak SM, Newton DL;
 XX
 DR WPI; 2000-365565/31.
 XX
 PT N-terminally modified RNase A targeted to and are cytotoxic to cancerous
 PT endothelial cells used to treat especially Kaposi's sarcoma.
 XX
 PS Example 9; Page 34; 51pp; English.
 XX
 CC The present sequence is a peptide which was inserted into expression
 CC vector pCANTAB5E to enable more flexible folding of the human eosinophil
 CC derived neurotoxin protein (EDN), which was expressed by the vector. This
 CC protein can be directed to cancerous cells using additional N-terminal
 CC peptides, where it exerts a cytotoxic effect. The protein can, therefore,
 CC be used to treat cancer, particularly Kaposi's sarcoma, and to
 CC selectively kill neoplastic and non-neoplastic endothelial cells
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 3; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 1 GGGGS 5

RESULT 12
 AAY54917
 ID AAY54917 standard; peptide; 5 AA.
 XX
 AC AAY54917;
 XX
 DT 14-FEB-2000 (first entry)
 XX

DE Linker from IL-12 fusion protein.
 XX
 KW Interleukin-12; IL-12; fusion protein; IL-12 p35 subunit; B7 protein;
 KW IL-12 p40 subunit; gene therapy; tumour; leukaemia; linker.
 XX
 OS Synthetic.
 XX
 PN US5994104-A.
 XX
 PD 30-NOV-1999.
 XX
 PF 08-NOV-1996; 96US-00751767.
 XX
 PR 08-NOV-1996; 96US-00751767.
 XX
 PA (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
 XX
 PI Anderson RJ, Prentice HG, Macdonald ID;
 XX
 DR WPI; 2000-038261/03.
 XX
 PT Nucleic acid constructs encoding interleukin-12 fusion proteins useful
 PT for treating leukemia and other cancers.
 XX
 PS Claim 3; Col 93; 73pp; English.
 XX
 CC This sequence represents a linker that can be used in an interleukin-12
 CC fusion protein. The invention relates to an isolated nucleic acid
 CC construct (I) comprising a region encoding an interleukin-12 (IL-12)
 CC fusion protein (comprising an IL-12 p35 subunit, an IL-12 p40 subunit and
 CC a linker peptide (joining the subunits)) and a region encoding a B7
 CC protein. (I) may be used to produce IL-12 fusion proteins according to
 CC standard recombinant DNA methodologies. The fusion proteins may be
 CC produced either in vitro in a fermentation culture or in vivo as part of
 CC a gene therapy protocol (in this case (I) is used to transform a patients
 CC cells, which then secrete the functional polypeptide to supplement the
 CC patients own production of IL-12 or to rectify mutations which lead to
 CC the expression of inactive polypeptides). The fusion proteins produced in
 CC this way may be used to treat any disease which responds to IL-12 such as
 CC tumours (both solid and dispersed of the kidney, breast, colon, ovarian
 CC and cervical tumours and melanomas) and in particular, tumours of the
 CC blood such as leukaemia. Alternatively, the polypeptides may be used as
 CC antigens in the production of antibodies to IL-12 and to assay for
 CC agonists and antagonists of its activity. The antibodies and antagonists
 CC may be used to inhibit the activity of IL-12. (I) may also be used
 CC diagnostically as a probe which hybridizes to sequences encoding IL-12
 CC and the antibodies may be used to detect the presence of IL-12
 CC polypeptides in samples. They may be used diagnostically to quantitate
 CC the expression of the polypeptide by patients and hence which subjects
 CC may be in need of restorative therapy
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 3; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 | | | |
Db 1 GGGGS 5

RESULT 13

AAY43750

ID AAY43750 standard; peptide; 5 AA.

XX

AC AAY43750;

XX

DT 11-FEB-2000 (first entry)

XX

DE Linker used to construct a bispecific single-chain antibody.

XX

KW bscCD19xCD3 antibody; bispecific single-chain fragment; CD19 antigen;

KW CD3 antigen; CD19-positive target cell; T-cell stimulation;

KW cytotoxic T-lymphocyte; B-cell malignancy; myasthenia gravis;

KW B-cell mediated autoimmune disease; Morbus Basedow;

KW Hashimoto thyroiditis; Goodpasture syndrome; B-cell depletion;

KW non-Hodgkin lymphoma; gene therapy; cancer; viral disease.

XX

OS Synthetic.

XX

PN WO9954440-A1.

XX

PD 28-OCT-1999.

XX

PF 21-APR-1999; 99WO-EP002693.

XX

PR 21-APR-1998; 98EP-00107269.

XX

PA (DOER/) DOERKEN B.

PA (RIET/) RIETHMUELLER G.

XX

PI Kufer P, Lutterbuese R, Bargou R, Loeffler A;

XX

DR WPI; 2000-013241/01.

XX

PT Novel multifunctional polypeptide for treating B-cell malignancies

PT especially non-Hodgkin lymphoma.

XX

PS Claim 10; Page 49; 91pp; English.

XX

CC The present sequence represents a linker used in the construction of
CC bispecific single-chain polypeptides of the invention. These polypeptides
CC comprise domains providing binding-site of immunoglobulin chains or
CC antibodies specifically recognizing CD19 and CD3 antigen. The polypeptide
CC destroys CD19-positive target cells without any need of T-cell pre and/or
CC co-stimulation, by recruiting cytotoxic T-lymphocytes and so specific
CC lysis by T-cells rather than a direct effect by an antibody is achieved.
CC The bispecific single-chain polypeptides, or nucleotides encoding them,
CC are used for the treatment of B-cell malignancies, B-cell mediated
CC autoimmune diseases like myasthenia gravis, Morbus Basedow, Hashimoto
CC thyroiditis or Goodpasture syndrome or for the depletion of B- cells and
CC more particularly non-Hodgkin lymphoma in mammals preferably human. They
CC can also delay the pathological conditions caused by these diseases, and

CC can be used for detecting these diseases. The polynucleotide is used for
CC gene therapy. The polypeptides are also used for identifying compounds
CC modulating B-cell/T-cell mediated immune response with can in turn be
CC used for treating cancer, its related diseases and also for inhibiting
CC viral diseases by preventing viral infection

XX

SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 3; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5

||||

Db 1 GGGGS 5

RESULT 14

AAB14535

ID AAB14535 standard; peptide; 5 AA.

XX

AC AAB14535;

XX

DT 24-NOV-2000 (first entry)

XX

DE Peptide linker for joining HIV-1 gp41 N- and C-terminal helices.

XX

KW HIV-1; gp41; N-helical domain; heptad repeat region; C-helical domain;

KW gp41 transmembrane-proximal amphipathic alpha-helical segment;

KW core 6-helix bundle; viral entry inhibition; immunogenic; antibody;

KW humoral response; broad spectrum vaccine; anti-HIV;

KW envelope glycoprotein; prophylaxis; therapy; peptide linker.

XX

OS Synthetic.

XX

PN WO200040616-A1.

XX

PD 13-JUL-2000.

XX

PF 10-JAN-2000; 2000WO-US000456.

XX

PR 08-JAN-1999; 99US-0115404P.

PR 07-JAN-2000; 2000US-00480336.

XX

PA (WILD/) WILD C T.

PA (WEIS/) WEISS C D.

XX

PI Wild CT, Weiss CD;

XX

DR WPI; 2000-465959/40.

XX

PT Raising neutralizing antibody response to human immunodeficiency virus,

PT comprises administering a polypeptide capable of forming a stable coiled-

PT coil solution structure.

XX

PS Disclosure; Page 15; 97pp; English.

XX

CC The invention relates to raising a neutralising antibody response to a
 CC broad spectrum of HIV (human immunodeficiency virus) strains and
 CC isolates, comprising the administration of a peptide which corresponds to
 CC or mimics highly conserved portions of the gp41 envelope glycoprotein
 CC which are important in mediating the process of viral entry into host
 CC cells. Such peptides can correspond to or mimic the coiled coil solution
 CC structure of the N-helical domain (the heptad repeat region), or can
 CC correspond or mimic the C-helical domain (the transmembrane-proximal
 CC amphipathic alpha-helical segment), or the gp41 core 6-helix bundle,
 CC which is formed by the interaction of the N- and C-helical domains of
 CC three gp41 proteins. The peptides can be administered either singly or as
 CC a combination (particularly a combination of N-helical and C-helical
 CC peptides), and can be multimerised. For example, N- and C-helical domain
 CC peptides can be alternately linked together to form a peptide which
 CC mimics the core 6-helix bundle. Administration of the peptide(s)
 CC generates a humoral response, with the production of antibodies against
 CC gp41 structures involved in viral entry. As these portions of gp41 are
 CC well conserved, such antibodies may be effective against a broad range of
 CC HIV strains and isolates. The peptide compositions may be administered as
 CC a prophylactic or therapeutic vaccine to generate antibodies which reduce
 CC or inhibit the ability of HIV to infect uninfected cells. A composition
 CC comprising polyclonal or monoclonal antibodies can be administered to
 CC reduce HIV infection of uninfected cells. Antibodies raised against entry
 CC -relevant gp41 structures may also be used therapeutically and as tools
 CC to further elucidate the mechanism of HIV cell entry. The present
 CC sequence represents a peptide linker which may be used to join peptides
 CC of the invention together to form multimers

XX

SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 3; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5

|||||

Db 1 GGGGS 5

RESULT 15

AAB00156

ID AAB00156 standard; peptide; 5 AA.

XX

AC AAB00156;

XX

DT 08-FEB-2001 (first entry)

XX

DE Linker used in sCD4-SCFv(17b) fusion protein.

XX

KW Fusion protein; HIV; human immunodeficiency virus; antibody; Fv; AIDS;

KW acquired immune deficiency syndrome; neutralisation; infection;

KW gene therapy; CD4; gp120; glycoprotein; resistance; vaccination;

KW binding domain; single chain antibody; chimera; chimeric protein.

XX

OS Synthetic.

XX

PN WO200055207-A1.

XX
 PD 21-SEP-2000.
 XX
 PF 16-MAR-2000; 2000WO-US006946.
 XX
 PR 16-MAR-1999; 99US-0124681P.
 XX
 PA (USSH) US NAT INST OF HEALTH.
 XX
 PI Berger EA, Del Castillo CM;
 XX
 DR WPI; 2000-638183/61.
 XX
 PT Novel neutralizing bispecific fusion proteins effective in viral such as
 PT HIV neutralization, comprises two different binding domains, inducing-
 PT binding domain and induced-binding domain functionally linked by linker.
 XX
 PS Claim 30; Page 45; 55pp; English.
 XX
 CC sCD4-SCFv(17b) is a neutralising bispecific fusion protein capable of
 CC binding to two sites of its target protein. The protein comprises a first
 CC binding domain capable of binding to an inducing site on the target
 CC protein, a second binding domain capable of forming neutralising complex
 CC with an induced epitope of the target protein and a linker connecting the
 CC binding domains. sCD4-SCFv(17b) comprises a soluble CD4 fragment
 CC (containing domains D1 and D2) fused to a single chain Fv portion of
 CC antibody 17b via a linker. sCD4-SCFv(17b), its variant, analogue or
 CC mimetic is used for inactivating gp120 protein of HIV, and for
 CC neutralising HIV. It is also used for blocking and preventing the binding
 CC of the viral or recombinant gp120 protein to soluble CD4 or lymphocyte
 CC CD4 and for inhibiting HIV replication. The chimeric proteins is
 CC therefore useful for treating HIV infection and also AIDS. It is are
 CC particularly useful in the prevention of infection during or immediately
 CC after HIV exposure (e.g., mother/infant transmission, post-exposure
 CC prophylaxis, and as a topical inhibitor) and for providing long term
 CC resistance to HIV infections and AIDS. Gene therapy is used to secrete
 CC the bispecific protein at mucosal surfaces, such as the vaginal, rectal
 CC or oral mucosa. The fusion proteins is highly potent, broadly cross-
 CC reactive with neutralising antibody with high in vivo activity and no Fc-
 CC mediated undesirable targetting properties. When the fusion protein is
 CC substantially derived from human proteins, it has minimal immunogenicity
 CC and toxicity in humans which is of great value in prevention of infection
 CC during or immediately after HIV exposure
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 28; DB 3; Length 5;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 1 GGGGS 5

Search completed: March 5, 2004, 16:22:50
 Job time : 6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:17:14 ; Search time 1.34259 Seconds
(without alignments)
192.262 Million cell updates/sec

Title: US-10-057-890A-16
Perfect score: 28
Sequence: 1 GGGGS 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	28	100.0	5	1	US-07-959-946-6	Sequence 6, Appli
2	28	100.0	5	1	US-08-176-500-140	Sequence 140, App
3	28	100.0	5	1	US-08-471-052A-140	Sequence 140, App
4	28	100.0	5	1	US-08-225-224-54	Sequence 54, Appl
5	28	100.0	5	1	US-08-236-918A-18	Sequence 18, Appl
6	28	100.0	5	1	US-08-463-163-1	Sequence 1, Appli
7	28	100.0	5	1	US-08-566-800A-58	Sequence 58, Appl
8	28	100.0	5	1	US-08-244-469-5	Sequence 5, Appli
9	28	100.0	5	1	US-08-189-331-140	Sequence 140, App
10	28	100.0	5	1	US-08-333-577-6	Sequence 6, Appli
11	28	100.0	5	1	US-08-575-361A-32	Sequence 32, Appl

12	28	100.0	5	2	US-08-564-955-64	Sequence 64, Appl
13	28	100.0	5	2	US-08-528-523-13	Sequence 13, Appl
14	28	100.0	5	2	US-08-537-874-62	Sequence 62, Appl
15	28	100.0	5	2	US-08-448-418-86	Sequence 86, Appl
16	28	100.0	5	2	US-08-471-939-140	Sequence 140, App
17	28	100.0	5	2	US-08-471-800-140	Sequence 140, App
18	28	100.0	5	2	US-08-932-589-58	Sequence 58, Appl
19	28	100.0	5	2	US-08-647-449-26	Sequence 26, Appl
20	28	100.0	5	2	US-08-821-840-1	Sequence 1, Appli
21	28	100.0	5	2	US-08-902-623-75	Sequence 75, Appl
22	28	100.0	5	2	US-08-471-068-140	Sequence 140, App
23	28	100.0	5	2	US-08-308-494A-1	Sequence 1, Appli
24	28	100.0	5	2	US-08-665-202-25	Sequence 25, Appl
25	28	100.0	5	2	US-08-809-668-12	Sequence 12, Appl
26	28	100.0	5	2	US-08-792-553-15	Sequence 15, Appl
27	28	100.0	5	2	US-08-751-767A-24	Sequence 24, Appl
28	28	100.0	5	2	US-08-818-253-40	Sequence 40, Appl
29	28	100.0	5	3	US-08-722-258-54	Sequence 54, Appl
30	28	100.0	5	3	US-09-397-951-12	Sequence 12, Appl
31	28	100.0	5	3	US-08-776-271-3	Sequence 3, Appli
32	28	100.0	5	3	US-08-621-859-64	Sequence 64, Appl
33	28	100.0	5	3	US-09-100-856A-62	Sequence 62, Appl
34	28	100.0	5	3	US-09-046-992-5	Sequence 5, Appli
35	28	100.0	5	3	US-09-215-035-3	Sequence 3, Appli
36	28	100.0	5	3	US-09-075-511-64	Sequence 64, Appl
37	28	100.0	5	3	US-09-099-015-64	Sequence 64, Appl
38	28	100.0	5	3	US-08-818-252-40	Sequence 40, Appl
39	28	100.0	5	3	US-09-091-814-1	Sequence 1, Appli
40	28	100.0	5	3	US-09-281-792B-26	Sequence 26, Appl
41	28	100.0	5	3	US-09-232-863-62	Sequence 62, Appl
42	28	100.0	5	3	US-09-362-805-1	Sequence 1, Appli
43	28	100.0	5	3	US-09-362-805-2	Sequence 2, Appli
44	28	100.0	5	3	US-09-367-953B-26	Sequence 26, Appl
45	28	100.0	5	3	US-09-133-508A-62	Sequence 62, Appl

ALIGNMENTS

RESULT 1

US-07-959-946-6

; Sequence 6, Application US/07959946

; Patent No. 5408038

; GENERAL INFORMATION:

; APPLICANT: Smith, Richard K.

; APPLICANT: Koduri, Raju

; APPLICANT: Young, Stephen G.

; APPLICANT: Witztum, Joseph L.

; APPLICANT: Curtiss, Linda K.

; TITLE OF INVENTION: Lipoprotein Assays Using Antibodies to a

; TITLE OF INVENTION: Pan Native Epitope and Recombinant Antigens

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker &

; ADDRESSEE: Milnamow, Ltd.

; STREET: 180 No. 5408038th Stetson, Suite 4700

; CITY: Chicago

```

; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/959,946
; FILING DATE: 19921008
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,706
; FILING DATE: 18-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)616-5400
; TELEFAX: (312)616-5460
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-959-946-6

```

```

Query Match          100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches      5; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 GGGGS 5
        |||||
Db      1 GGGGS 5

```

RESULT 2

US-08-176-500-140

```

; Sequence 140, Application US/08176500
; Patent No. 5498538
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 141
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/176,500
;      FILING DATE:
;      CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/013,416
;      FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  1101-143
;  TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  140:
;      SEQUENCE CHARACTERISTICS:
;          LENGTH:  5 amino acids
;          TYPE:  amino acid
;          STRANDEDNESS:  single
;          TOPOLOGY:  unknown
;      MOLECULE TYPE:  peptide
US-08-176-500-140

```

```

Query Match          100.0%;  Score 28;  DB 1;  Length 5;
Best Local Similarity 100.0%;  Pred. No. 3e+05;
Matches      5;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

```

```

Qy      1 GGGGS 5
        |||||
Db      1 GGGGS 5

```

RESULT 3

US-08-471-052A-140

```

; Sequence 140, Application US/08471052A
; Patent No. 5625033
;  GENERAL INFORMATION:
;      APPLICANT:  Kay, B. K.
;      APPLICANT:  Fowlkes, D. M.
;      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
;      NUMBER OF SEQUENCES:  166
;      CORRESPONDENCE ADDRESS:
;          ADDRESSEE:  Pennie & Edmonds
;          STREET:  1155 Avenue of the Americas
;          CITY:  New York
;          STATE:  New York
;          COUNTRY:  U.S.A.
;          ZIP:  10036-2711
;  COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,052A
; FILING DATE: 06-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 140:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-471-052A-140

```

```

Query Match          100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches      5; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 GGGGS 5
        |||||
Db      1 GGGGS 5

```

RESULT 4

```

US-08-225-224-54
; Sequence 54, Application US/08225224
; Patent No. 5635599
; GENERAL INFORMATION:
; APPLICANT: PASTAN, Ira
; APPLICANT: KREITMAN, Robert J.
; TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND
; TITLE OF INVENTION: CIRCULARLY PERMUTED FUSION PROTEINS
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,224
; FILING DATE: 8-APR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-225-224-54

Query Match 100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
| | | |
Db 1 GGGGS 5

RESULT 5

US-08-236-918A-18

; Sequence 18, Application US/08236918A

; Patent No. 5674704

; GENERAL INFORMATION:

; APPLICANT: Alderson, Mark R.

; APPLICANT: Goodwin, Raymond G.

; APPLICANT: Smith, Craig A.

; TITLE OF INVENTION: Cytokine Designated 4-lBB Ligand

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kathryn A. Anderson, Immunex Corporation

; STREET: 51 University Street

; CITY: Seattle

; STATE: Washington

; COUNTRY: US

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Power Macintosh

; OPERATING SYSTEM: Apple 7.5.3

; SOFTWARE: Microsoft Word, Version #6.0.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/236,918A

; FILING DATE: 06-May-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/060,843

; FILING DATE: 07-May-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Anderson, Kathryn A.

; REGISTRATION NUMBER: 32,172

; REFERENCE/DOCKET NUMBER: 2801-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-236-918A-18

Query Match 100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 6

US-08-463-163-1

; Sequence 1, Application US/08463163
; Patent No. 5696237
; GENERAL INFORMATION:
; APPLICANT: FitzGerald, David J.
; APPLICANT: Chaudhary, Vijay K.
; APPLICANT: Pastan, Ira H.
; APPLICANT: Waldmann, Thomas A.
; APPLICANT: Queen, Cary L.
; TITLE OF INVENTION: Recombinant Antibody-Toxin Fusion Protein
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,163
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/227,227
; FILING DATE: 22-JAN-1981
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/911,227
; FILING DATE: 24-SEP-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,361

```

; FILING DATE: 21-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/865,722
; FILING DATE: 08-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 015280-12211
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-463-163-1

```

```

Query Match          100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches      5; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 GGGGS 5
        |||||
Db      1 GGGGS 5

```

RESULT 7

```

US-08-566-800A-58
; Sequence 58, Application US/08566800A
; Patent No. 5736364
; GENERAL INFORMATION:
; APPLICANT: Kelley, Robert F.
; APPLICANT: Lazarus, Robert A.
; APPLICANT: Lee, Geoffrey F.
; TITLE OF INVENTION: No. 5736364e1 Factor VIIa Inhibitors
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/566,800A
; FILING DATE: 04-Dec-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kubinec, Jeffrey S.

```

; REGISTRATION NUMBER: 36,575
; REFERENCE/DOCKET NUMBER: P0958B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8228
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-566-800A-58

Query Match 100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 8

US-08-244-469-5

; Sequence 5, Application US/08244469
; Patent No. 5736387
; GENERAL INFORMATION:
; APPLICANT: Paul, Ralph W.
; APPLICANT: Overell, Robert
; TITLE OF INVENTION: ENVELOPE FUSION VECTORS FOR USE IN GENE
; TITLE OF INVENTION: DELIVERY
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,469
; FILING DATE: 01-JUN-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Dylan, Tyler M.
; REGISTRATION NUMBER: 37,612
; REFERENCE/DOCKET NUMBER: 22627-20007.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141 MRSNFOERS SFO
; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-244-469-5

Query Match 100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 9

US-08-189-331-140

; Sequence 140, Application US/08189331
; Patent No. 5747334
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 186
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/189,331
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 140:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-189-331-140

Query Match 100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 10

US-08-333-577-6

; Sequence 6, Application US/08333577
; Patent No. 5786206
; GENERAL INFORMATION:
; APPLICANT: Smith, Richard K.
; APPLICANT: Koduri, Raju
; APPLICANT: Young, Stephen G.
; APPLICANT: Witztum, Joseph L.
; APPLICANT: Curtiss, Linda K.
; TITLE OF INVENTION: Lipoprotein Assays Using Antibodies to a
; TITLE OF INVENTION: Pan Native Epitope and Recombinant Antigens
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Shore, Sutker &
; ADDRESSEE: Milnamow, Ltd.
; STREET: 180 No. 5786206th Stetson, Suite 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/333,577
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; REFERENCE/DOCKET NUMBER: SCRF 234.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)616-5400
; TELEFAX: (312)616-5460
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-333-577-6

Query Match 100.0%; Score 28; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||

Db 1 GGGGS 5

RESULT 11

US-08-575-361A-32

; Sequence 32, Application US/08575361A

; Patent No. 5792640

; GENERAL INFORMATION:

; APPLICANT: Chandrasegaran, Srinivasan

; TITLE OF INVENTION: A GENERAL METHOD TO CLONE HYBRID

; TITLE OF INVENTION: RESTRICTION ENDONUCLEASES USING lig GENE

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cushman Darby & Cushman L.L.P.

; STREET: 1100 New York Avenue, NW, Ninth Floor, East

; STREET: Tower

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005-3918

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/575,361A

; FILING DATE: 20-DEC-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Kokulis, Paul N.

; REGISTRATION NUMBER: 16,773

; REFERENCE/DOCKET NUMBER: PNK/4130/213779/DJP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-861-3000

; TELEFAX: 202-822-0944

; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 32:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

US-08-575-361A-32

Query Match 100.0%; Score 28; DB 1; Length 5;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||

Db 1 GGGGS 5

RESULT 12
 US-08-564-955-64
 ; Sequence 64, Application US/08564955
 ; Patent No. 5811238
 ; GENERAL INFORMATION:
 ; APPLICANT: STEMMER, WILLEM P.C.
 ; APPLICANT: CRAMERI, ANDREAS M.
 ; TITLE OF INVENTION: METHODS FOR GENERATING POLYNUCLEOTIDES
 ; TITLE OF INVENTION: HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION
 AND
 ; TITLE OF INVENTION: RECOMBINATION
 ; NUMBER OF SEQUENCES: 67
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW
 ; STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
 ; CITY: SAN FRANCISCO
 ; STATE: CALIFORNIA
 ; COUNTRY: U.S.A.
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/564,955
 ; FILING DATE: 30-NOV-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/198,431
 ; FILING DATE: 17-FEB-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/537,874
 ; FILING DATE: 30-OCT-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/02126
 ; FILING DATE: 17-FEB-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DUNN, TRACY J.
 ; REGISTRATION NUMBER: 34,587
 ; REFERENCE/DOCKET NUMBER: 16528J-014611US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 326-2400
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 64:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 US-08-564-955-64

Query Match 100.0%; Score 28; DB 2; Length 5;
 Best Local Similarity 100.0%; Pred. No. 3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5

Db |||||
 1 GGGGS 5

RESULT 13

US-08-528-523-13

; Sequence 13, Application US/08528523
; Patent No. 5824782
; GENERAL INFORMATION:
; APPLICANT: Hoelzer, Wolfgang
; APPLICANT: von Hoegen, Ilka
; APPLICANT: Strittmatter, Wolfgang
; APPLICANT: Matzku, Siegfried
; TITLE OF INVENTION: Immunoconjugates II
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Boulevard, Suite 1400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/528,523
; FILING DATE: 06-NOV-1992
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94114572.4
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hamlet-King, Diana
; REGISTRATION NUMBER: 33,302
; REFERENCE/DOCKET NUMBER: Merck 1717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-243-6333
; TELEFAX: 703-243-6410
; TELEX: 64191
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
US-08-528-523-13

Query Match 100.0%; Score 28; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 14

US-08-537-874-62

; Sequence 62, Application US/08537874

; Patent No. 5830721

; GENERAL INFORMATION:

; APPLICANT: Stemmer, Willem P.C.

; APPLICANT: Cramer, Andreas

; TITLE OF INVENTION: DNA Mutagenesis by Random Fragmentation

; TITLE OF INVENTION: and Reassembly

; NUMBER OF SEQUENCES: 62

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/537,874

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP PCT/US95/02126

; FILING DATE: 17-FEB-1995

; APPLICATION NUMBER: US 08/198,431

; FILING DATE: 17-FEB-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Liebeschuetz, Joe

; REGISTRATION NUMBER: 37,505

; REFERENCE/DOCKET NUMBER: 018097-014610

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 62:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

US-08-537-874-62

Query Match 100.0%; Score 28; DB 2; Length 5;

Best Local Similarity 100.0%; Pred. No. 3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||

RESULT 15

US-08-448-418-86

; Sequence 86, Application US/08448418

; Patent No. 5837242

; GENERAL INFORMATION:

; APPLICANT: Holliger, Kaspar-Philipp

; APPLICANT: Griffiths, Andrew D

; APPLICANT: Hoogenboom, Hendricus RJM

; APPLICANT: Malmqvist, Magnus

; APPLICANT: Marks, James D

; APPLICANT: McGuinness, Brian T

; APPLICANT: Pope, Anthony R

; APPLICANT: Prospero, Terence D

; APPLICANT: Winter, Gregory P

; TITLE OF INVENTION: Multivalent and Multispecific Binding

; TITLE OF INVENTION: Proteins, Their Manufacture and Use

; NUMBER OF SEQUENCES: 106

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall O'Toole Gerstein Murray and Borun

; STREET: 6300 Sears Tower 233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/448,418

; FILING DATE: 14-MAY-1996

; CLASSIFICATION: 435

; CLASSIFICATION: C12N 15/62, 15/70, C07K 1/00

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB93/02492

; FILING DATE: 03-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9225453.1

; FILING DATE: 04-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9300816.7

; FILING DATE: 16-JAN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 93303614.7

; FILING DATE: 10-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9319969.3

; FILING DATE: 22-SEP-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: David W. Clough

; REGISTRATION NUMBER: 36,107

; REFERENCE/DOCKET NUMBER: 28111/32651

; INFORMATION FOR SEQ ID NO: 86:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide linker
US-08-448-418-86

Query Match 100.0%; Score 28; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

Search completed: March 5, 2004, 16:30:38
Job time : 2.34259 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:16:19 ; Search time 1.14198 Seconds
(without alignments)
421.163 Million cell updates/sec

Title: US-10-057-890A-16
Perfect score: 28
Sequence: 1 GGGGS 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pirl:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	28	100.0	37	2	S29113		dipteridin homolog
2	28	100.0	64	2	A86333		hypothetical prote
3	28	100.0	66	2	H84489		hypothetical prote
4	28	100.0	69	1	MIEC77		microcin B17 precu
5	28	100.0	78	2	E84686		hypothetical prote
6	28	100.0	80	2	T10550		hypothetical prote
7	28	100.0	81	2	PC2047		grain-softness pro
8	28	100.0	82	2	S19774		glycine-rich prote
9	28	100.0	85	2	T32664		hypothetical prote
10	28	100.0	92	2	PQ0743		grain-softness pro
11	28	100.0	97	2	T48330		hypothetical prote
12	28	100.0	100	2	T49621		hypothetical prote
13	28	100.0	102	2	T25332		hypothetical prote

14	28	100.0	104	2	T02612	hypothetical prote
15	28	100.0	104	2	JC4190	holotricin 3 precu
16	28	100.0	108	2	G86252	hypothetical prote
17	28	100.0	109	2	S58673	RNA-binding protei
18	28	100.0	110	2	AC2391	RNA-binding protei
19	28	100.0	114	2	S28821	transcription fact
20	28	100.0	115	2	T35387	hypothetical prote
21	28	100.0	119	2	T07695	hypothetical prote
22	28	100.0	120	2	A81109	hypothetical prote
23	28	100.0	120	2	D83415	hypothetical prote
24	28	100.0	122	2	T04118	mitochondrial proc
25	28	100.0	122	2	D86754	prophage pi2 prote
26	28	100.0	125	2	T16247	hypothetical prote
27	28	100.0	128	2	T30428	hypothetical prote
28	28	100.0	131	2	H69062	molybdenum transpo
29	28	100.0	133	2	G75432	hypothetical prote
30	28	100.0	135	2	S55647	hypothetical prote
31	28	100.0	136	2	T02870	globulin 2 precurs
32	28	100.0	136	2	T29282	hypothetical prote
33	28	100.0	139	2	C87544	hypothetical prote
34	28	100.0	140	2	AC3088	hypothetical prote
35	28	100.0	144	2	S04069	glycine-rich prote
36	28	100.0	144	2	S35716	glycine-rich prote
37	28	100.0	144	2	T34730	probable gas vesic
38	28	100.0	145	1	JQ1062	glycine-rich prote
39	28	100.0	145	2	E84469	probable glycine-r
40	28	100.0	148	2	S46514	puroindoline-b pre
41	28	100.0	148	2	I38881	caudal-type homeot
42	28	100.0	149	2	T23179	hypothetical prote
43	28	100.0	150	2	C86224	hypothetical prote
44	28	100.0	152	2	T04811	STIG1 protein homo
45	28	100.0	155	2	C86206	hypothetical prote

ALIGNMENTS

RESULT 1

S29113

dipteracin homolog - flesh fly (*Sarcophaga peregrina*)

C;Species: *Sarcophaga peregrina*

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-May-1999

C;Accession: S29113

R;Ishikawa, M.; Kubo, T.; Natori, S.

Biochem. J. 287, 573-578, 1992

A;Title: Purification and characterization of a dipteracin homologue from *Sarcophaga peregrina* (flesh fly).

A;Reference number: S29113; MUID:93074996; PMID:1445217

A;Accession: S29113

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-37 <ISH>

Query Match 100.0%; Score 28; DB 2; Length 37;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGS 5
|||||
Db 19 GGGGS 23

RESULT 2

A86333

hypothetical protein T20H2.25 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C;Accession: A86333

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: A86333

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-64 <STO>

A;Cross-references: GB:AE005172; NID:g8779001; PIDN:AAF79916.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

Query Match 100.0%; Score 28; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGS 5
|||||
Db 25 GGGGS 29

RESULT 3

H84489

hypothetical protein At2g10020 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: H84489

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams,

M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: H84489

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-66 <STO>

A;Cross-references: GB:AE002093; NID:g4558680; PIDN:AAD22697.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g10020

A;Map position: 2

Query Match 100.0%; Score 28; DB 2; Length 66;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 22 GGGGS 26

RESULT 4

MIEC77

microcin B17 precursor - *Escherichia coli* plasmid pMccB17

C;Species: *Escherichia coli*

C;Date: 30-Jun-1988 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999

C;Accession: A25219; A32058; I41099; A58368; S67977

R;Davagnino, J.; Herrero, M.; Furlong, D.; Moreno, F.; Kolter, R.

Proteins 1, 230-238, 1986

A;Title: The DNA replication inhibitor microcin B17 is a forty-three-amino-acid protein containing sixty percent glycine.

A;Reference number: A25219; MUID:88217867; PMID:3329729

A;Accession: A25219

A;Molecule type: DNA

A;Residues: 1-69 <DAV>

A;Cross-references: GB:M15469; NID:g146787; PIDN:AAA24141.1; PID:g146788

R;Genilloud, O.; Moreno, F.; Kolter, R.

J. Bacteriol. 171, 1126-1135, 1989

A;Title: DNA sequence, products, and transcriptional pattern of the genes involved in production of the DNA replication inhibitor microcin B17.

A;Reference number: A32058; MUID:89123111; PMID:2644225

A;Accession: A32058

A;Molecule type: DNA

A;Residues: 1-69 <GEN>

A;Cross-references: GB:M24253; NID:g341145; PIDN:AAA72741.1; PID:g522290

R;Connell, N.; Han, Z.; Moreno, F.; Kolter, R.

Mol. Microbiol. 1, 195-201, 1987

A;Title: An *E. coli* promoter induced by the cessation of growth.

A;Reference number: I41099; MUID:88216163; PMID:2835580

A;Accession: I41099

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-14 <CON>

A;Cross-references: EMBL:X06417; NID:g41978; PIDN:CAA29725.1; PID:g41979
R;Li, Y.M.; Milne, J.C.; Madison, L.L.; Kolter, R.; Walsh, C.T.
Science 274, 1188-1193, 1996
A;Title: From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase.
A;Reference number: A58368; MUID:97053605; PMID:8895467
A;Accession: A58368
A;Molecule type: protein
A;Residues: 27-38 <LIY>
A;Experimental source: Escherichia coli strain ZK4(pY113)
A;Note: mass spectroscopy of peptides and biosynthetic intermediates
R;Yorgey, P.; Lee, J.; Koerdel, J.; Vivas, E.; Warner, P.; Jebaratnam, D.; Kolter, R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4519-4523, 1994
A;Title: Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor.
A;Reference number: A58375; MUID:94240167; PMID:8183941
A;Contents: annotation; (1)H-NMR spectroscopy of modified peptides
R;Bayer, A.; Freund, S.; Jung, G.
Eur. J. Biochem. 234, 414-426, 1995
A;Title: Post-translational heterocyclic backbone modifications in the 43-peptide antibiotic microcin B17. Structure elucidation and NMR study of a (13)C, (15)N-labelled gyrase inhibitor.
A;Reference number: S67977; MUID:96128168; PMID:8536683
A;Accession: S67977
A;Status: preliminary
A;Molecule type: protein
A;Residues: 27-38 <BAY>
C;Genetics:
A;Gene: mcbA
A;Genome: plasmid pMccB17
C;Function:
A;Description: inhibits DNA gyrase, stopping DNA replication
A;Note: active against a large number of gram-negative enteric bacteria
C;Superfamily: microcin
C;Keywords: antibiotic; DNA replication inhibitor; oxazole/thiazole ring
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-69/Product: microcin B17 #status experimental <MAT>
F;39-40/Cross-link: oxazole (Gly-Ser) #status experimental
F;40-41/Cross-link: thiazole (Ser-Cys) #status experimental
F;47-48/Cross-link: thiazole (Gly-Cys) #status experimental
F;50-51/Cross-link: thiazole (Gly-Cys) #status experimental
F;54-55/Cross-link: thiazole (Gly-Cys) #status experimental
F;55-56/Cross-link: oxazole (Cys-Ser) #status experimental
F;61-62/Cross-link: oxazole (Gly-Ser) #status experimental
F;64-65/Cross-link: oxazole (Gly-Ser) #status experimental

Query Match 100.0%; Score 28; DB 1; Length 69;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGS 5
|||||
Db 36 GGGGS 40

RESULT 5

E84686

hypothetical protein At2g28570 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: E84686

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: E84686

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-78 <STO>

A;Cross-references: GB:AE002093; NID:g4510404; PIDN:AAD21491.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g28570

A;Map position: 2

Query Match 100.0%; Score 28; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 53 GGGGS 57

RESULT 6

T10550

hypothetical protein T12G13.70 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999

C;Accession: T10550

R;Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C. submitted to the Protein Sequence Database, June 1999

A;Reference number: Z16533

A;Accession: T10550

A;Molecule type: DNA

A;Residues: 1-80 <BEV>

A;Cross-references: EMBL:AL080252; GSPDB:GN00062; ATSP:T12G13.70

A;Experimental source: cultivar Columbia; BAC clone T12G13

C;Genetics:

A;Gene: ATSP:T12G13.70

A;Map position: 4

Query Match 100.0%; Score 28; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
Db 27 GGGGS 31

RESULT 7

PC2047

grain-softness protein - wheat (fragments)

C;Species: Triticum aestivum (common wheat)

C;Date: 05-Aug-1994 #sequence_revision 05-Aug-1994 #text_change 14-Sep-1994

C;Accession: PC2047

R;Jolly, C.J.; Rahman, S.; Kortt, A.A.; Higgins, T.J.V.

Theor. Appl. Genet. 86, 589-597, 1993

A;Title: Characterisation of the wheat Mr 15000 grain-softness protein and analysis of the relationship between its accumulation in the whole seed and grain softness.

A;Reference number: PQ0743

A;Accession: PC2047

A;Molecule type: protein

A;Residues: 1-18;19-24;25-31;32-38;39-45;46-51;52-56;57-60;61-65;66-71;72-77;78-81 <JOL>

C;Comment: This protein is the product of the Ha locus and thus be the major factor that determines the milling characteristics of bread wheats.

C;Keywords: seed

Query Match 100.0%; Score 28; DB 2; Length 81;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
Db 4 GGGGS 8

RESULT 8

S19774

glycine-rich protein - tomato (fragment)

C;Species: Lycopersicon esculentum (tomato)

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 23-Jul-1999

C;Accession: S19774

R;Parsons, B.L.

submitted to the EMBL Data Library, May 1991

A;Reference number: S19773

A;Accession: S19774

A;Molecule type: mRNA

A;Residues: 1-82 <PAR>

A;Cross-references: EMBL:X59883; NID:g19321; PIDN:CAA42538.1; PID:g19322

C;Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology

Query Match 100.0%; Score 28; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
Db 72 GGGGS 76

RESULT 9

T32664

hypothetical protein F16B4.7 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Jun-2000

C;Accession: T32664

R;Davidson, S.; Wohldmann, P.; Bauer, C.; O'Neal, D.

submitted to the EMBL Data Library, December 1997

A;Description: The sequence of *C. elegans* cosmid F16B4.

A;Reference number: Z21208

A;Accession: T32664

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-85 <DAV>

A;Cross-references: EMBL:AF039048; PIDN:AAB94238.1; GSPDB:GN00023; CESP:F16B4.7

A;Experimental source: strain Bristol N2; clone F16B4

C;Genetics:

A;Gene: CESP:F16B4.7

A;Map position: 5

A;Introns: 36/1

C;Superfamily: Arabidopsis glycine-rich protein 3

Query Match 100.0%; Score 28; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|
Db 39 GGGGS 43

RESULT 10

PQ0743

grain-softness protein - wheat (fragments)

C;Species: *Triticum aestivum* (common wheat)

C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 23-Mar-1995

C;Accession: PQ0743

R;Jolly, C.J.; Rahman, S.; Kortt, A.A.; Higgins, T.J.V.

Theor. Appl. Genet. 86, 589-597, 1993

A;Title: Characterisation of the wheat Mr 15000 grain-softness protein and analysis of the relationship between its accumulation in the whole seed and grain softness.

A;Reference number: PQ0743

A;Accession: PQ0743

A;Molecule type: protein

A;Residues: 1-92 <JOL>

A;Experimental source: seed

C;Keywords: seed

Query Match 100.0%; Score 28; DB 2; Length 92;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|
Db 39 GGGGS 43

Db

4 GGGGS 8

RESULT 11

T48330

hypothetical protein F15A17.120 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C;Accession: T48330

R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De Keyser, A.; Neyt, P.; Rouze, P.; Van Den Daele, H.; Villaroel, R.; Gielen, J.; Van Montagu, M.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24491

A;Accession: T48330

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-97 <BEV>

A;Cross-references: EMBL:AL163002

A;Experimental source: cultivar Columbia; BAC clone F15A17

C;Genetics:

A;Map position: 5

A;Introns: 7/1

A;Note: F15A17.120

Query Match 100.0%; Score 28; DB 2; Length 97;

Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5

||||

Db 47 GGGGS 51

RESULT 12

T49621

hypothetical protein B5O22.30 [imported] - Neurospora crassa

C;Species: Neurospora crassa

C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C;Accession: T49621

R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, G.; Mewes, H.W.; Mannhaupt, G.

submitted to the Protein Sequence Database, May 2000

A;Reference number: Z25022

A;Accession: T49621

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-100 <SCH>

A;Cross-references: EMBL:AL355932; GSPDB:GN00116; NCSP:B5O22.30

A;Experimental source: BAC clone B5O22; strain OR74A

C;Genetics:

A;Gene: NCSP:B5O22.30

A;Map position: 6

A;Introns: 22/1; 52/1

Query Match 100.0%; Score 28; DB 2; Length 100;

Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 11 GGGGS 15

RESULT 13

T25332

hypothetical protein T26H5.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C;Accession: T25332

R;Gardner, A.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z20017

A;Accession: T25332

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-102 <WIL>

A;Cross-references: EMBL:Z82056; PIDN:CAB04855.1; GSPDB:GN00023; CESP:T26H5.4

A;Experimental source: clone T26H5

C;Genetics:

A;Gene: CESP:T26H5.4

A;Map position: 5

A;Introns: 13/1; 96/1

C;Superfamily: hypothetical protein K01D12.8

Query Match 100.0%; Score 28; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 52 GGGGS 56

RESULT 14

T02612

hypothetical protein At2g26120 [imported] - *Arabidopsis thaliana*

N;Alternate names: hypothetical protein T19L18.7

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001

C;Accession: T02612; F84656

R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.;
Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.;
Venter, J.C.

submitted to the EMBL Data Library, August 1998

A;Description: *Arabidopsis thaliana* chromosome II BAC T19L18 genomic sequence.

A;Reference number: Z14681

A;Accession: T02612

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-104 <ROU>

A;Cross-references: EMBL:AC004747; NID:g3413696; PID:g3413702

A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: F84656

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-104 <STO>

A;Cross-references: GB:AE002093; NID:g3413702; PIDN:AAC31225.1; GSPDB:GN00139

C;Genetics:

A;Gene: T19L18.7; At2g26120

A;Map position: 2

A;Introns: 49/3

Query Match	100.0%;	Score 28;	DB 2;	Length 104;
Best Local Similarity	100.0%;	Pred. No. 3.2e+02;		
Matches	5;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

Qy	1	GGGGS	5
Db	96	GGGGS	100

RESULT 15

JC4190

holotricin 3 precursor - *Holotrichia diomphalia*

N;Alternate names: antifungal protein

C;Species: *Holotrichia diomphalia*

C;Date: 04-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999

C;Accession: JC4190

R;Lee, S.Y.; Moon, H.J.; Kurata, S.; Natori, S.; Lee, B.L.

Biol. Pharm. Bull. 18, 1049-1052, 1995

A;Title: Purification and cDNA cloning of an antifungal protein from the hemolymph of *Holotrichia diomphalia* larvae.

A;Reference number: JC4190; MUID:96073722; PMID:8535393

A;Accession: JC4190

A;Molecule type: mRNA

A;Residues: 1-104 <LEE>

A;Cross-references: DDBJ:D13744; NID:g1088433; PIDN:BAA02889.1; PID:d1003394; PID:g1786168

C;Comment: This protein is a Gly- and His-rich protein and a constitutive protein of larval hemolymph.

C;Keywords: hemolymph

F;1-20/Domain: signal sequence #status predicted <SIG>

F;21-104/Product: holotricin 3 #status predicted <MAT>

Query Match	100.0%;	Score 28;	DB 2;	Length 104;
Best Local Similarity	100.0%;	Pred. No. 3.2e+02;		
Matches	5;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
Db 69 GGGGS 73

Search completed: March 5, 2004, 16:28:57
Job time : 2.14198 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:22:54 ; Search time 2.70062 Seconds
(without alignments)
390.935 Million cell updates/sec

Title: US-10-057-890A-16
Perfect score: 28
Sequence: 1 GGGGS 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	ID	Description

1	28	100.0	5	9	US-09-287-849-44	Sequence 44, Appl
2	28	100.0	5	9	US-09-147-142-31	Sequence 31, Appl
3	28	100.0	5	9	US-09-214-645-1	Sequence 1, Appli
4	28	100.0	5	9	US-09-858-616-2	Sequence 2, Appli
5	28	100.0	5	9	US-09-779-233-45	Sequence 45, Appl
6	28	100.0	5	9	US-09-989-789-3	Sequence 3, Appli
7	28	100.0	5	9	US-09-976-787-21	Sequence 21, Appl
8	28	100.0	5	9	US-09-192-854-180	Sequence 180, App
9	28	100.0	5	9	US-09-761-962-36	Sequence 36, Appl
10	28	100.0	5	9	US-09-333-527-5	Sequence 5, Appli
11	28	100.0	5	9	US-09-925-796-8	Sequence 8, Appli
12	28	100.0	5	9	US-09-815-837-116	Sequence 116, App
13	28	100.0	5	9	US-09-033-525-5	Sequence 5, Appli
14	28	100.0	5	9	US-09-779-451-7	Sequence 7, Appli
15	28	100.0	5	9	US-09-941-450-8	Sequence 8, Appli
16	28	100.0	5	9	US-09-818-247-25	Sequence 25, Appl
17	28	100.0	5	9	US-09-883-777-10	Sequence 10, Appl
18	28	100.0	5	9	US-09-867-262-3	Sequence 3, Appli
19	28	100.0	5	9	US-09-780-933-22	Sequence 22, Appl
20	28	100.0	5	9	US-09-480-236-10	Sequence 10, Appl
21	28	100.0	5	9	US-09-731-558-6	Sequence 6, Appli
22	28	100.0	5	9	US-09-828-708-123	Sequence 123, App
23	28	100.0	5	9	US-09-885-551A-3	Sequence 3, Appli
24	28	100.0	5	9	US-09-756-283A-14	Sequence 14, Appl
25	28	100.0	5	9	US-09-144-886-4	Sequence 4, Appli
26	28	100.0	5	9	US-09-999-745-56	Sequence 56, Appl
27	28	100.0	5	9	US-09-942-087A-8	Sequence 8, Appli
28	28	100.0	5	9	US-09-942-090-8	Sequence 8, Appli
29	28	100.0	5	9	US-09-554-000-40	Sequence 40, Appl
30	28	100.0	5	9	US-09-792-793A-1	Sequence 1, Appli
31	28	100.0	5	9	US-09-792-793A-2	Sequence 2, Appli
32	28	100.0	5	10	US-09-846-033B-212	Sequence 212, App
33	28	100.0	5	10	US-09-990-186-3	Sequence 3, Appli
34	28	100.0	5	10	US-09-897-844-8	Sequence 8, Appli
35	28	100.0	5	10	US-09-989-994-3	Sequence 3, Appli
36	28	100.0	5	10	US-09-911-261A-23	Sequence 23, Appl
37	28	100.0	5	10	US-09-942-024-84	Sequence 84, Appl
38	28	100.0	5	10	US-09-942-098-84	Sequence 84, Appl
39	28	100.0	5	10	US-09-969-748C-38	Sequence 38, Appl
40	28	100.0	5	10	US-09-992-124A-61	Sequence 61, Appl
41	28	100.0	5	10	US-09-949-039-37	Sequence 37, Appl
42	28	100.0	5	13	US-10-087-426-3	Sequence 3, Appli
43	28	100.0	5	13	US-10-057-505-15	Sequence 15, Appl
44	28	100.0	5	13	US-10-115-984-6	Sequence 6, Appli
45	28	100.0	5	13	US-10-153-159-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
 US-09-287-849-44
 ; Sequence 44, Application US/09287849
 ; Patent No. US20020009459A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.

```

; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible
; OTHER INFORMATION: polylinker
US-09-287-849-44

```

```

Query Match          100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches      5; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 GGGGS 5
        |||||
Db      1 GGGGS 5

```

RESULT 2

US-09-147-142-31

```

; Sequence 31, Application US/09147142
; Patent No. US20020018749A1
; GENERAL INFORMATION:
; APPLICANT: HUDSON, Peter John
; APPLICANT: KORTT, Alex Andrew
; APPLICANT: IRVING, Robert Alexander
; APPLICANT: ATWELL, John Leslie
; TITLE OF INVENTION: HIGH AVIDITY POLYVALENT AND POLYSPECIFIC REAGENTS
; FILE REFERENCE: 016786/0212
; CURRENT APPLICATION NUMBER: US/09/147,142
; CURRENT FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: PCT/AU98/00212
; EARLIER FILING DATE: 1998-03-26
; EARLIER APPLICATION NUMBER: AU PO 5917
; EARLIER FILING DATE: 1997-03-27

```

; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide linker
US-09-147-142-31

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 3

US-09-214-645-1

; Sequence 1, Application US/09214645
; Patent No. US20020028443A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: METHOD OF DNA SHUFFLING WITH
; TITLE OF INVENTION: POLYNUCLEOTIDES PRODUCED BY BLOCKING OR INTERRUPTING A
; TITLE OF INVENTION: SYNTHESIS OR AMPLIFICATION PROCESS
; FILE REFERENCE: DIVER1220-2
; CURRENT APPLICATION NUMBER: US/09/214,645
; CURRENT FILING DATE: 1999-09-27
; PRIOR APPLICATION NUMBER: PCT/US97/12239
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: linker peptide
US-09-214-645-1

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 4

US-09-858-616-2

; Sequence 2, Application US/09858616
; Patent No. US20020031771A1
; GENERAL INFORMATION:


```

; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; TITLE OF INVENTION: SEQUENCE BASED SCREENING
; FILE REFERENCE: DIVER1210-6
; CURRENT APPLICATION NUMBER: US/09/858,616
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US 09/571,499
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 09/557,276
; PRIOR FILING DATE: 2000-04-24
; PRIOR APPLICATION NUMBER: US 08/692,002
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: US 60/008,317
; PRIOR FILING DATE: 1995-12-07
; PRIOR APPLICATION NUMBER: US 08/944,795
; PRIOR FILING DATE: 1997-10-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Linker peptide
US-09-858-616-2

```

```

Query Match          100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches      5; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      1 GGGGS 5
        |||||
Db      1 GGGGS 5

```

RESULT 5

US-09-779-233-45

```

; Sequence 45, Application US/09779233
; Patent No. US20020045158A1
; GENERAL INFORMATION:
; APPLICANT: Case, Casey
; TITLE OF INVENTION: CELLS FOR DRUG DISCOVERY
; FILE REFERENCE: 8325-0010
; CURRENT APPLICATION NUMBER: US/09/779,233
; CURRENT FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: linker
US-09-779-233-45

```

```

Query Match          100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;

```

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||

Db 1 GGGGS 5

RESULT 6

US-09-989-789-3

; Sequence 3, Application US/09989789
; Patent No. US20020063379A1
; GENERAL INFORMATION:
; APPLICANT: LIU, Qiang
; TITLE OF INVENTION: POSITION DEPENDENT RECOGNITION OF GNN NUCLEOTIDE
; TITLE OF INVENTION: TRIPLETS BY ZINC FINGERS
; FILE REFERENCE: 8325-0011.20 / S11-US2
; CURRENT APPLICATION NUMBER: US/09/989,789
; CURRENT FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 4085
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide linker
US-09-989-789-3

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||

Db 1 GGGGS 5

RESULT 7

US-09-976-787-21

; Sequence 21, Application US/09976787
; Patent No. US20020064528A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Zhenping
; APPLICANT: Witte, Larry
; TITLE OF INVENTION: Antibodies Specific to KDR and Uses Thereof
; FILE REFERENCE: 11245/46505
; CURRENT APPLICATION NUMBER: US/09/976,787
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 09/493,539
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 60/117,726
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: WordPerfect 8.0 for Windows
; SEQ ID NO 21
; LENGTH: 5
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptide linker
US-09-976-787-21

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 8

US-09-192-854-180

; Sequence 180, Application US/09192854
; Patent No. US20020068276A1
; GENERAL INFORMATION:
; APPLICANT: Winter, Greg
; APPLICANT: Tomlinson, Ian
; TITLE OF INVENTION: Methods for Selecting Functional Peptides
; FILE REFERENCE: 3789/72916
; CURRENT APPLICATION NUMBER: US/09/192,854
; CURRENT FILING DATE: 1998-11-17
; EARLIER APPLICATION NUMBER: 60/066,729
; EARLIER FILING DATE: 1997-11-21
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 180
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Linker peptide
; OTHER INFORMATION: for connecting variable domains.
US-09-192-854-180

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 9

US-09-761-962-36

; Sequence 36, Application US/09761962
; Patent No. US20020077285A1
; GENERAL INFORMATION:
; APPLICANT: Memorial Sloan-Kettering Cancer Center
; TITLE OF INVENTION: Identification and Characterization of Multiple Splice
; TITLE OF INVENTION: Variants of Mu-
; TITLE OF INVENTION: opioid Receptor (MOR-1) Gene
; FILE REFERENCE: 830002-2000.1

; CURRENT APPLICATION NUMBER: US/09/761,962
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/743,872
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: basic unit of a linking peptide
US-09-761-962-36

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|
|
|
|
Db 1 GGGGS 5

RESULT 10

US-09-333-527-5

; Sequence 5, Application US/09333527
; Patent No. US20020078472A1
; GENERAL INFORMATION:
; APPLICANT: Paul CHRISTOU; Eva STROGER; Rainer FISCHER; Carmen MARTIN-
VAQUERO; Stefan
; TITLE OF INVENTION: METHODS AND MEANS FOR EXPRESSION OF MAMMALIAN
POLYPEPTIDES
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,527
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/089,322
; FILING DATE: June 15, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: KL/JIC 202.1 - JEL
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 318-3000
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5
; TYPE: amino acid
; TOPOLOGY: linear
US-09-333-527-5

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 11

US-09-925-796-8

; Sequence 8, Application US/09925796
; Patent No. US20020081614A1
; GENERAL INFORMATION:
; APPLICANT: Case, Casey C.
; APPLICANT: Zhang, Lei
; APPLICANT: Sangamo Biosciences, Inc.
; TITLE OF INVENTION: Functional Genomics Using Zinc Finger Proteins
; FILE REFERENCE: 019496-002000US
; CURRENT APPLICATION NUMBER: US/09/925,796
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/395,448
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: 09/229,037
; PRIOR FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:linker
US-09-925-796-8

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 12

US-09-815-837-116

; Sequence 116, Application US/09815837
; Patent No. US20020082411A1

```
; GENERAL INFORMATION:
; APPLICANT: Carter, Darrick
; APPLICANT: Zhu, Shirley
; APPLICANT: Arimilli, Subhashini
; APPLICANT: Wang, Aijun
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Immune Medators and Related Methods
; FILE REFERENCE: 014058-005670US
; CURRENT APPLICATION NUMBER: US/09/815,837
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,274
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,249
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/264,003
; PRIOR FILING DATE: 2001-01-23
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 116
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:downstream
; OTHER INFORMATION: linker for CO596
US-09-815-837-116
```

```
Query Match          100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches      5; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
```

```
Qy      1 GGGGS 5
        |||||
Db      1 GGGGS 5
```

RESULT 13

US-09-033-525-5

```
; Sequence 5, Application US/09033525
; Patent No. US20020090374A1
; GENERAL INFORMATION:
; APPLICANT: Yarkoni, Shai
; APPLICANT: Ben-Yehudah, Ahmi
; APPLICANT: Azar, Yehudith
; APPLICANT: Aqeilan, Rami
; APPLICANT: Belotstotsky, Ruth
; APPLICANT: Lorberboum-Galski, Haya
; TITLE OF INVENTION: CHIMERIC PROTEINS WITH CELL-TARGETING
; TITLE OF INVENTION: SPECIFICITY AND APOPTOSIS-INDUCING ACTIVITIES
; FILE REFERENCE: 9457-009-999
; CURRENT APPLICATION NUMBER: US/09/033,525
; CURRENT FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 5
; TYPE: PRT
```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Flexible polylinker
US-09-033-525-5

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 14

US-09-779-451-7

; Sequence 7, Application US/09779451
; Patent No. US20020094521A1
; GENERAL INFORMATION:
; APPLICANT: Wild, Carl T.
; APPLICANT: Allaway, Graham P.
; TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors
; FILE REFERENCE: 1900.0300003
; CURRENT APPLICATION NUMBER: US/09/779,451
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/235,901
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/181,543
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: REPEAT
; LOCATION: (1)..(5)
; OTHER INFORMATION: (GGGGS)x, where x is 1, 2, 3, 4, or 5
; NAME/KEY: misc_feature
; OTHER INFORMATION: Preferred amino acid residues
US-09-779-451-7

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

RESULT 15

US-09-941-450-8

; Sequence 8, Application US/09941450
; Patent No. US20020094529A1
; GENERAL INFORMATION:

; APPLICANT: Case, Casey C.
; APPLICANT: Urnov, Fyodor
; TITLE OF INVENTION: GENE IDENTIFICATION
; FILE REFERENCE: S7.US3 / 8325-0007.20
; CURRENT APPLICATION NUMBER: US/09/941,450
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: 09/395,448
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:linker
US-09-941-450-8

Query Match 100.0%; Score 28; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 7.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 1 GGGGS 5

Search completed: March 5, 2004, 16:33:44
Job time : 2.70062 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:15:44 ; Search time 3.3179 Seconds
(without alignments)
475.479 Million cell updates/sec

Title: US-10-057-890A-16
Perfect score: 28
Sequence: 1 GGGGS 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description

1	28	100.0	17	2	Q9R4Y9	Q9r4y9 alcaligenes
2	28	100.0	17	11	Q64450	Q64450 mus musculu
3	28	100.0	20	2	Q9R582	Q9r582 neisseria m
4	28	100.0	20	10	Q8W238	Q8w238 glycine max
5	28	100.0	23	4	Q9UC00	Q9uc00 homo sapien
6	28	100.0	26	10	Q42226	Q42226 arabidopsis
7	28	100.0	31	2	Q9F1I5	Q9f1i5 enterococcu
8	28	100.0	37	5	Q9TWW2	Q9tww2 sarcophaga
9	28	100.0	40	4	Q13833	Q13833 homo sapien
10	28	100.0	40	4	Q13832	Q13832 homo sapien
11	28	100.0	45	10	Q7XMY7	Q7xmy7 oryza sativ
12	28	100.0	50	10	Q943L9	Q943l9 oryza sativ
13	28	100.0	50	10	Q84YS5	Q84ys5 oryza sativ
14	28	100.0	50	16	Q98C82	Q98c82 rhizobium l
15	28	100.0	53	10	Q8LNH3	Q8lnh3 oryza sativ
16	28	100.0	60	10	Q8H647	Q8h647 oryza sativ
17	28	100.0	61	16	Q82GM4	Q82gm4 streptomyce
18	28	100.0	62	4	Q8N2Z2	Q8n2z2 homo sapien
19	28	100.0	63	10	Q94E89	Q94e89 oryza sativ
20	28	100.0	63	10	Q7XNX4	Q7xnx4 oryza sativ
21	28	100.0	63	12	Q8VB65	Q8vb65 white spot
22	28	100.0	64	10	Q9LNS5	Q9lns5 arabidopsis
23	28	100.0	65	10	Q94J57	Q94j57 oryza sativ
24	28	100.0	66	5	Q9XZQ8	Q9xzzq8 caenorhabdi
25	28	100.0	66	6	Q9XT04	Q9xt04 macropus ru
26	28	100.0	66	10	Q9SK19	Q9sk19 arabidopsis
27	28	100.0	66	10	Q8H3Q4	Q8h3q4 oryza sativ
28	28	100.0	67	10	Q42070	Q42070 arabidopsis
29	28	100.0	67	10	Q8L3R6	Q8l3r6 oryza sativ
30	28	100.0	67	10	Q84YV0	Q84yv0 oryza sativ
31	28	100.0	69	10	Q8S2H8	Q8s2h8 oryza sativ
32	28	100.0	70	4	Q7Z6C2	Q7z6c2 homo sapien
33	28	100.0	72	10	Q84ST5	Q84st5 oryza sativ
34	28	100.0	74	10	Q9LLN8	Q9lln8 oryza sativ
35	28	100.0	76	12	Q8VAF7	Q8vaf7 white spot
36	28	100.0	78	10	Q9SK01	Q9sk01 arabidopsis
37	28	100.0	80	10	Q9SUF7	Q9suf7 arabidopsis
38	28	100.0	81	10	Q8L4N9	Q8l4n9 oryza sativ
39	28	100.0	81	10	Q7XWD8	Q7xwd8 oryza sativ
40	28	100.0	81	16	Q89TN1	Q89tn1 bradyrhizob
41	28	100.0	82	10	Q04130	Q04130 lycopersico
42	28	100.0	82	10	Q8H799	Q8h799 arabidopsis
43	28	100.0	83	10	Q9M7I4	Q9m7i4 zea mays (m
44	28	100.0	84	10	Q8H5X1	Q8h5x1 oryza sativ
45	28	100.0	84	12	Q98ZK1	Q98zk1 human parvo

ALIGNMENTS

RESULT 1

Q9R4Y9

ID	Q9R4Y9	PRELIMINARY;	PRT;	17 AA.
AC	Q9R4Y9;			
DT	01-MAY-2000	(TrEMBLrel. 13, Created)		
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)		
DT	01-JUN-2000	(TrEMBLrel. 14, Last annotation update)		

DE Aromatic amine dehydrogenase beta subunit (Fragment).
 OS Alcaligenes faecalis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Alcaligenes.
 OX NCBI_TaxID=511;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=94245619; PubMed=8188594;
 RA Govindaraj S., Eisenstein E., Jones L.H., Sanders-Loehr J.,
 RA Chistoserdov A.Y., Davidson V.L., Edwards S.L.;
 RT "Aromatic amine dehydrogenase, a second tryptophan tryptophylquinone
 RT enzyme.";
 RL J. Bacteriol. 176:2922-2929(1994).
 SQ SEQUENCE 17 AA; 1510 MW; 6EEEEAB9D89D2661 CRC64;

Query Match 100.0%; Score 28; DB 2; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 2 GGGGS 6

RESULT 2

Q64450

ID Q64450 PRELIMINARY; PRT; 17 AA.
 AC Q64450;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Uridine kinase (EC 2.7.1.48) (Fragment).
 GN UMPK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Ropp P.A., Traut T.W.;
 RT "Cloning and expression of a cDNA encoding uridine kinase from
 RT mouse.";
 RL Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; U57332; AAB01998.1; -.
 DR MGD; MGI:98904; Umpk.
 DR GO; GO:0016301; F:kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0004849; F:uridine kinase activity; IEA.
 KW Kinase; Transferase.
 FT NON_TER 17 17
 SQ SEQUENCE 17 AA; 1464 MW; 14E427CBA1168634 CRC64;

Query Match 100.0%; Score 28; DB 11; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 5 GGGGS 9

RESULT 3

Q9R582

ID Q9R582 PRELIMINARY; PRT; 20 AA.
AC Q9R582;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Transferrin-binding protein 2 (Fragment).
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE.
RX MEDLINE=93307625; PubMed=8319886;
RA Griffiths E., Stevenson P., Byfield P., Ala'Aldeen D.A.,
RA Borriello S.P., Holland J., Parsons T., Williams P.;
RT "Antigenic relationships of transferrin-binding proteins from
RT Neisseria meningitidis, N. gonorrhoeae and Haemophilus influenzae:
RT cross-reactivity of antibodies to NH2-terminal peptides.";
RL FEMS Microbiol. Lett. 109:85-91(1993).
SQ SEQUENCE 20 AA; 1977 MW; 6000EE169F09227E CRC64;

Query Match 100.0%; Score 28; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 3 GGGGS 7

RESULT 4

Q8W238

ID Q8W238 PRELIMINARY; PRT; 20 AA.
AC Q8W238;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE GT-2 factor (Fragment).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21471140; PubMed=11587508;
RA O'Grady K., Goekjian V.H., Naim C.J., Nagao R.T., Key J.L.;
RT "The transcript abundance of GmGT-2, a new member of the GT-2 family
RT of transcription factors from soybean, is down-regulated by light in a
RT phytochrome-dependent manner.";

RL Plant Mol. Biol. 47:367-378(2001).
DR EMBL; AF372500; AAL65126.1; -.
FT NON_TER 20 20
SQ SEQUENCE 20 AA; 1692 MW; F65C75CD9C6B663B CRC64;

Query Match 100.0%; Score 28; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 11 GGGGS 15

RESULT 5

Q9UC00

ID Q9UC00 PRELIMINARY; PRT; 23 AA.
AC Q9UC00;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Enhancement of wound HEALING process.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130623; PubMed=7829572;
RA Pierschbacher M.D., Polarek J.W., Craig W.S., Tschopp J.F.,
RA Sipes N.J., Harper J.R.;
RL J. Cell. Biochem. 56:150-154(1994).
DR GO; GO:0009611; P:response to wounding; TAS.
SQ SEQUENCE 23 AA; 2268 MW; CE73999CB9903891 CRC64;

Query Match 100.0%; Score 28; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 11 GGGGS 15

RESULT 6

Q42226

ID Q42226 PRELIMINARY; PRT; 26 AA.
AC Q42226;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Seed maturation protein (Fragment).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Seed;
 RA Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
 RL Submitted (FEB-1994) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; Z29850; CAA82818.1; -.
 FT NON_TER 1 1
 SQ SEQUENCE 26 AA; 2370 MW; 6E0902E39464466A CRC64;

Query Match 100.0%; Score 28; DB 10; Length 26;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 11 GGGGS 15

RESULT 7

Q9F1I5

ID Q9F1I5 PRELIMINARY; PRT; 31 AA.
 AC Q9F1I5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN EP0010.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OG Plasmid pAM373.
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20453452; PubMed=10998166;
 RA De Boever E.H., Clewell D.B., Fraser C.M.;
 RT "Enterococcus faecalis conjugative plasmid pAM373: complete nucleotide
 RT sequence and genetic analyses of sex pheromone response."
 RL Mol. Microbiol. 37:1327-1341(2000).
 DR EMBL; AE002565; AAG40421.1; -.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 31 AA; 3509 MW; 4E19CB94B3DB9421 CRC64;

Query Match 100.0%; Score 28; DB 2; Length 31;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 18 GGGGS 22

RESULT 8

Q9TWW2

ID Q9TWW2 PRELIMINARY; PRT; 37 AA.
 AC Q9TWW2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Dipteracin homolog (Fragment).
 OS Sarcophaga peregrina (Flesh fly) (Boettcherisca peregrina).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
 OC Sarcophagidae; Sarcophaga.
 OX NCBI_TaxID=7386;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=LARVAL HEMOLYMPH;
 RX MEDLINE=93074996; PubMed=1445217;
 RA Ishikawa M., Kubo T., Natori S.;
 RT "Purification and characterization of a dipteracin homologue from
 RT Sarcophaga peregrina (flesh fly).";
 RL Biochem. J. 287:573-578(1992).
 CC -!- FUNCTION: BACTERICIDAL ACTIVITY AGAINST GRAM-NEGATIVE BACTERIA
 CC E.COLI AND S.SONNEI.
 CC -!- TISSUE SPECIFICITY: SYNTHESIZED BY THE FAT BODY AND IS SECRETED
 CC INTO THE HEMOLYMPH.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSION IN THE LARVAE STARTS A FEW MINUTES
 CC AFTER THE INJURY OF THE BODY WALL REACHING A MAXIMUM AFTER ABOUT
 CC 10 HOURS. THE MAXIMUM LASTS FOR AT LEAST 3 HOURS.
 CC -!- INDUCTION: IN RESPONSE TO INJURY OF THE BODY WALL OF THE LARVAE.
 DR PIR; S29113; S29113.
 DR GO; GO:0006952; P:defense response; IEA.
 DR GO; GO:0006805; P:xenobiotic metabolism; IEA.
 KW Insect immunity; Antibiotic.
 FT DOMAIN 18 22 POLY-GLY.
 FT NON_TER 37 37
 SQ SEQUENCE 37 AA; 3928 MW; E3BAC8105D2DABC7 CRC64;

Query Match 100.0%; Score 28; DB 5; Length 37;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 19 GGGGS 23

RESULT 9

Q13833

ID Q13833 PRELIMINARY; PRT; 40 AA.
 AC Q13833;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE B2 bradykinin receptor basal promoter, allele BP-58-T (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96209920; PubMed=8655154;
 RA Braun A., Maier E., Kammerer S., Mueller B., Roscher A.A.;

RT "A novel sequence polymorphism in the promoter region of the human
 RT bradykinin B2-receptor gene.";
 RL Hum. Genet. 97:688-689(1996).
 DR EMBL; X91664; CAA62852.1; -.
 DR GO; GO:0004872; F:receptor activity; IEA.
 KW Receptor.
 FT NON_TER 1 1
 FT VARIANT 18 19 IT -> XS.
 FT NON_TER 40 40
 SQ SEQUENCE 40 AA; 4152 MW; 1408E9AD371EE17F CRC64;

Query Match 100.0%; Score 28; DB 4; Length 40;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 6 GGGGS 10

RESULT 10

Q13832

ID Q13832 PRELIMINARY; PRT; 40 AA.
 AC Q13832;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE B2 bradykinin receptor basal promoter, allele BP-58-C (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96209920; PubMed=8655154;
 RA Braun A., Maier E., Kammerer S., Mueller B., Roscher A.A.;
 RT "A novel sequence polymorphism in the promoter region of the human
 RT bradykinin B2-receptor gene.";
 RL Hum. Genet. 97:688-689(1996).
 DR EMBL; X91663; CAA62851.1; -.
 DR GO; GO:0004872; F:receptor activity; IEA.
 KW Receptor.
 FT NON_TER 1 1
 FT VARIANT 18 19 TT -> XS.
 FT NON_TER 40 40
 SQ SEQUENCE 40 AA; 4140 MW; 3908E9AD371EF4A5 CRC64;

Query Match 100.0%; Score 28; DB 4; Length 40;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 6 GGGGS 10

RESULT 11

Q7XMY7

ID Q7XMY7 PRELIMINARY; PRT; 45 AA.
AC Q7XMY7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNBa0027G07.9 protein.
GN OSJNBa0027G07.9.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,
RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,
RA Chen L., Fan D.L., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,
RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,
RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
RA Zhang R.Q., Guan J.P., Hong G.F.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL662937; CAE04373.1; -.
SQ SEQUENCE 45 AA; 4397 MW; 206F05C2B215436F CRC64;

Query Match 100.0%; Score 28; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 6 GGGGS 10

RESULT 12

Q943L9

ID Q943L9 PRELIMINARY; PRT; 50 AA.
AC Q943L9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE P0031D11.15 protein.
GN P0031D11.15.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0031D11.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP003231; BAB67882.1; -.
DR Gramene; Q943L9; -.
SQ SEQUENCE 50 AA; 4759 MW; 686CAE4584F62907 CRC64;

Query Match 100.0%; Score 28; DB 10; Length 50;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 4 GGGGS 8

RESULT 13

Q84YS5

ID Q84YS5 PRELIMINARY; PRT; 50 AA.
AC Q84YS5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE OSJNBa0027N13.2 protein.
GN OSJNBa0027N13.2.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
RT clone:OSJNBa0027N13.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP005641; BAC57367.1; -.
SQ SEQUENCE 50 AA; 5184 MW; D7797E195F50E5E2 CRC64;

Query Match 100.0%; Score 28; DB 10; Length 50;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 4 GGGGS 8

RESULT 14

Q98C82

ID Q98C82 PRELIMINARY; PRT; 50 AA.
AC Q98C82;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein msl5259.
GN MSL5259.
OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003006; BAB51739.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 50 AA; 5260 MW; CE20ED738F5286D0 CRC64;

Query Match 100.0%; Score 28; DB 16; Length 50;
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 27 GGGGS 31

RESULT 15

Q8LNH3

ID Q8LNH3 PRELIMINARY; PRT; 53 AA.
 AC Q8LNH3;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN OSJNBA0078001.18.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Kim M.M.,
 RA Overton II L.L., Bera J.J., Tsitrin T., Krol M.I., Jarrahi B.B.,
 RA Jin S.S., Koo H., Zismann V., Hsiao J., Blunt S., Vanaken S.S.,
 RA Utterback T.T., Feldblyum T.V., Yang Q.Q., Haas B.J., Suh B.B.,
 RA Peterson J.J., Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBA0078001 genomic sequence.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA The Rice Chromosome 10 Sequencing Consortium;
 RT "In-depth view of structure, activity, and evolution of rice

RT chromosome 10.";
RL Science 300:1566-1569(2003).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC079888; AAM93673.1; -.
DR EMBL; AE017109; AAP54462.1; -.
DR Gramene; Q8LNH3; -.
KW Hypothetical protein.
SQ SEQUENCE 53 AA; 5493 MW; 9440B5801C3650B6 CRC64;

Query Match 100.0%; Score 28; DB 10; Length 53;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
| | | | |
Db 20 GGGGS 24

Search completed: March 5, 2004, 16:27:32
Job time : 4.3179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 5, 2004, 16:15:14 ; Search time 0.679012 Seconds
(without alignments)
383.426 Million cell updates/sec

Title: US-10-057-890A-16
Perfect score: 28
Sequence: 1 GGGGS 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	28	100.0	17	1	PPK5_PERAM	P82617 periplaneta
2	28	100.0	69	1	MCBA_ECOLI	P05834 escherichia
3	28	100.0	82	1	NUMM_MOUSE	P52503 mus musculu
4	28	100.0	104	1	HOL3_HOLDI	Q25055 holotrichia
5	28	100.0	110	1	FUS1_HUMAN	O75896 homo sapien
6	28	100.0	115	1	YOB9_STRCO	Q9xai3 streptomyce
7	28	100.0	118	1	CRYP_CRYPA	P52753 cryphonectr
8	28	100.0	144	1	GRP1_CHERU	P11898 chenopodium
9	28	100.0	144	1	GRP9_DAUCA	P37703 daucus caro
10	28	100.0	144	1	GVA1_STRCO	Q9zcl3 streptomyce
11	28	100.0	148	1	PUIB_WHEAT	Q10464 triticum ae
12	28	100.0	157	1	GRP_DAUCA	Q03878 daucus caro
13	28	100.0	163	1	CGC8_MOUSE	Q9d187 mus musculu
14	28	100.0	164	1	SSB_BRAJA	Q89150 bradyrhizob
15	28	100.0	165	1	GRP1_ORYSA	P25074 oryza sativ
16	28	100.0	165	1	SSB_MYCSM	Q9afi5 mycobacteri
17	28	100.0	168	1	HY5_ARATH	O24646 arabidopsis

18	28	100.0	172	1	SSB_RHILO	Q98m41 rhizobium l
19	28	100.0	176	1	GRP7_ARATH	Q03250 arabidopsis
20	28	100.0	179	1	CBFA_MAIZE	P25209 zea mays (m
21	28	100.0	182	1	K2C3_BOVIN	P04261 bos taurus
22	28	100.0	183	1	OLEC_BRANA	P29526 brassica na
23	28	100.0	187	1	SCX_CHICK	P59101 gallus gall
24	28	100.0	190	1	XYN_TRIHA	P48793 trichoderma
25	28	100.0	192	1	VG49_BPMD2	O64239 mycobacteri
26	28	100.0	201	1	GR2B_ARATH	Q38896 arabidopsis
27	28	100.0	201	1	TWS1_PANTR	Q8mi03 pan troglod
28	28	100.0	202	1	TWS1_HUMAN	Q15672 homo sapien
29	28	100.0	204	1	TWS1_HYLCO	Q8mie7 hylobates c
30	28	100.0	205	1	CME1_BACSU	P39694 bacillus su
31	28	100.0	206	1	TWS1_MOUSE	P26687 mus musculu
32	28	100.0	209	1	IM23_HUMAN	O14925 homo sapien
33	28	100.0	209	1	IM23_RAT	O35093 rattus norv
34	28	100.0	214	1	GRP2_NICSY	P27484 nicotiana s
35	28	100.0	222	1	ERF4_ARATH	O80340 arabidopsis
36	28	100.0	225	1	RNT_PSESM	Q87xm0 pseudomonas
37	28	100.0	226	1	SP25_CANFA	Q28250 canis famil
38	28	100.0	226	1	SP25_MOUSE	Q9cyn2 mus musculu
39	28	100.0	231	1	RNH_STRCO	Q9x7r6 streptomyce
40	28	100.0	232	1	GS28_DROME	Q9ve50 drosophila
41	28	100.0	235	1	BRT1_RAT	P55007 rattus norv
42	28	100.0	236	1	RTN3_HUMAN	O95197 homo sapien
43	28	100.0	237	1	RTN3_MOUSE	Q9es97 mus musculu
44	28	100.0	238	1	FLGH_BRUME	Q8yb19 brucella me
45	28	100.0	238	1	FLGH_BRUSU	Q8fxc2 brucella su

ALIGNMENTS

RESULT 1

PPK5_PERAM

ID PPK5_PERAM STANDARD; PRT; 17 AA.
AC P82617;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pyrokinin-5 (Pea-PK-5) (FXPRL-amide).
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blattoidea;
OC Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
RC TISSUE=Abdominal perisymphathetic organs;
RX MEDLINE=99212469; PubMed=10196736;
RA Predel R., Kellner R., Nachman R.J., Holman G.M., Rapus J., Gaede G.;
RT "Differential distribution of pyrokinin-isoforms in cerebral and
RT abdominal neurohemal organs of the American cockroach.";
RL Insect Biochem. Mol. Biol. 29:139-144(1999).
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=20189894; PubMed=10723010;

RA Predel R., Eckert M.;
 RT "Tagma-specific distribution of FXPRLamides in the nervous system of
 RT the American cockroach.";
 RL J. Comp. Neurol. 419:352-363(2000).
 CC -!- FUNCTION: Mediates visceral muscle contractile activity (myotropic
 CC activity).
 CC -!- TISSUE SPECIFICITY: Mainly in abdominal perisymphathetic organs and
 CC to a lesser extent in retrocerebral complex.
 CC -!- MASS SPECTROMETRY: MW=1651.7; METHOD=MALDI.
 CC -!- SIMILARITY: Belongs to the pyrokinin family.
 DR InterPro; IPR001484; Pyrokinin.
 DR PROSITE; PS00539; PYROKININ; 1.
 KW Neuropeptide; Amidation; Pyrokinin.
 FT MOD_RES 17 17 AMIDATION.
 SQ SEQUENCE 17 AA; 1653 MW; 8527162EA45BBA54 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 17;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 1 GGGGS 5

RESULT 2

MCBA_ECOLI

ID MCBA_ECOLI STANDARD; PRT; 69 AA.
 AC P05834;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bacteriocin microcin B17 precursor (MCB17).
 GN MCBA.
 OS Escherichia coli.
 OG Plasmid IncFII pMccB17.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88217867; PubMed=3329729;
 RA Davagnino J., Herrero M., Furlong D., Moreno F., Kolter R.;
 RT "The DNA replication inhibitor microcin B17 is a
 RT forty-three-amino-acid protein containing sixty percent glycine.";
 RL Proteins 1:230-238(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89123111; PubMed=2644225;
 RA Genilloud O., Moreno F., Kolter R.;
 RT "DNA sequence, products, and transcriptional pattern of the genes
 RT involved in production of the DNA replication inhibitor microcin
 RT B17.";
 RL J. Bacteriol. 171:1126-1135(1989).
 RN [3]
 RP SEQUENCE OF 1-14 FROM N.A.
 RX MEDLINE=88216163; PubMed=2835580;

RA Conell N., Han Z., Moreno F., Kolter R.;
 RT "An E. coli promoter induced by the cessation of growth.";
 RL Mol. Microbiol. 1:195-201(1987).
 RN [4]
 RP PARTIAL SEQUENCE OF 27-69.
 RA Bayer A., Stevanovic S., Freund S., Metzger J.W., Jung G.;
 RT "Isolation and structure elucidation of the 43-peptide antibiotic
 RT microcin B17.";
 RL (In) Schneider C.H., Eberles A.N. (eds.);
 RL Peptides 1992, pp.117-118, Escom Science Publishers, Leiden (1993).
 RN [5]
 RP FUNCTION.
 RX MEDLINE=91122055; PubMed=1846808;
 RA Vizan J.L., Hernandez-Chico C., del Castillo I., Moreno F.;
 RT "The peptide antibiotic microcin B17 induces double-strand cleavage
 RT of DNA mediated by E. coli DNA gyrase.";
 RL EMBO J. 10:467-476(1991).
 RN [6]
 RP STRUCTURE BY NMR OF 1-26.
 RX MEDLINE=98213789; PubMed=9545435;
 RA Roy R.S., Kim S., Baleja J.D., Walsh C.T.;
 RT "Role of the microcin B17 propeptide in substrate recognition:
 RT solution structure and mutational analysis of McbA1-26.";
 RL Chem. Biol. 5:217-228(1998).
 CC -!- FUNCTION: THIS GLYCINE-RICH PEPTIDE ANTIBIOTIC INHIBITS DNA
 CC REPLICATION IN MANY ENTERIC BACTERIA, THAT LEADS TO INDUCTION OF
 CC THE SOS REPAIR SYSTEM, MASSIVE DNA DEGRADATION AND CELL DEATH.
 CC B17 INHIBITS TYPE II TOPOISOMERASE BY TRAPPING AN ENZYME - DNA
 CC CLEAVABLE COMPLEX.
 CC -!- PTM: The processed N-terminus does not resemble a typical
 CC secretion signal sequence.
 CC -!- PTM: THE CYS RESIDUES AS WELL AS SOME GLY AND CYS ARE POST-
 CC TRANSLATIONALLY MODIFIED. MODIFICATIONS INCLUDE THE FORMATION OF
 CC FOUR THIAZOLE AND FOUR OXAZOLE RINGS THAT RESULT, RESPECTIVELY,
 CC FROM THE CONDENSATION OF FOUR SERINE SIDE CHAINS WITH THE CARBONYL
 CC GROUP OF THE PRECEDING AMINO ACID.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M15469; AAA24141.1; -.
 DR EMBL; M24253; AAA72741.1; -.
 DR EMBL; X06417; CAA29725.1; -.
 DR PIR; A25219; MIEC77.
 DR PDB; 2MLP; 22-JUL-98.
 KW DNA replication inhibitor; Antibiotic; Bacteriocin; Plasmid;
 KW 3D-structure.
 FT PROPEP 1 26
 FT CHAIN 27 69 BACTERIOICIN MICROICIN B17.
 FT DOMAIN 26 39 POLY-GLY.
 FT CROSSLNK 39 40 Oxazole (Gly-Ser).
 FT CROSSLNK 40 41 Thiazole (Ser-Cys).

FT	CROSSLNK	47	48	Thiazole (Gly-Cys).
FT	CROSSLNK	50	51	Thiazole (Gly-Cys).
FT	CROSSLNK	54	55	Thiazole (Gly-Cys).
FT	CROSSLNK	55	56	Oxazole (Cys-Ser).
FT	CROSSLNK	61	62	Oxazole (Gly-Ser).
FT	CROSSLNK	64	65	Oxazole (Gly-Ser).
FT	TURN	6	7	
FT	HELIX	8	15	
FT	TURN	16	16	
FT	HELIX	18	21	
SQ	SEQUENCE	69 AA;	6013 MW;	0B1D159A832638A8 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 69;
 Best Local Similarity 100.0%; Pred. No. 98;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 36 GGGGS 40

RESULT 3

NUMM_MOUSE

ID NUMM_MOUSE STANDARD; PRT; 82 AA.
 AC P52503;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE NADH-ubiquinone oxidoreductase 13 kDa-A subunit (EC 1.6.5.3)
 DE (EC 1.6.99.3) (Complex I-13KD-A) (CI-13KD-A) (Fragment).
 GN NDUFS6 OR IP13.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1;
 RX MEDLINE=95331630; PubMed=7607554;
 RA Watson J.D., Beckett-Jones B., Roy R.N., Green N.C., Flynn T.G.;
 RT "Genomic sequence, structural organization and evolutionary
 RT conservation of the 13.2-kDa subunit of rat NADH:ubiquinone
 RT oxidoreductase."
 RL Gene 158:275-280(1995).
 CC -!- FUNCTION: Transfer of electrons from NADH to the respiratory
 CC chain. The immediate electron acceptor for the enzyme is believed
 CC to be ubiquinone. This is a component of the iron-sulfur (IP)
 CC fragment of the enzyme.
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
 CC -!- SUBUNIT: Mammalian complex I is composed of 45 different subunits.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane; matrix side.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; L38438; AAB64010.1; -.
DR MGD; MGI:107932; Ndufs6.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
FT NON_TER 1 1
FT NON_TER 82 82
SQ SEQUENCE 82 AA; 9330 MW; C923FFE245A9BD27 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|
Db 50 GGGGS 54

RESULT 4

HOL3_HOLDI

ID HOL3_HOLDI STANDARD; PRT; 104 AA.
AC Q25055;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Holotricin 3 precursor.
OS Holotrichia diomphalia.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Scarabaeiformia;
OC Scarabaeidae; Melolonthinae; Holotrichia.
OX NCBI_TaxID=33394;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-40.
RC TISSUE=Larval hemolymph;
RX MEDLINE=96073722; PubMed=8535393;
RA Lee S.Y., Moon H.J., Kurata S., Natori S., Lee B.L.;
RT "Purification and cDNA cloning of an antifungal protein from the
RT hemolymph of Holotrichia diomphalia larvae."
RL Biol. Pharm. Bull. 18:1049-1052(1995).
CC -!- FUNCTION: Has antifungal activity against C.albicans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: TO TENECIN 3.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D13744; BAA02889.1; -.
DR PIR; JC4190; JC4190.
KW Insect immunity; Antibiotic; Hemolymph; Fungicide; Signal; Repeat.
FT SIGNAL 1 20

FT	CHAIN	21	104	HOLOTRICIN 3.
FT	DOMAIN	27	98	18 X 4 AA APPROXIMATE TANDEM REPEATS OF
FT				H-G-G-G.
FT	REPEAT	27	30	1.
FT	REPEAT	31	34	2.
FT	REPEAT	35	38	3.
FT	REPEAT	39	42	4.
FT	REPEAT	43	46	5.
FT	REPEAT	47	50	6.
FT	REPEAT	51	54	7.
FT	REPEAT	55	58	8.
FT	REPEAT	59	62	9.
FT	REPEAT	63	66	10.
FT	REPEAT	67	70	11.
FT	REPEAT	71	74	12.
FT	REPEAT	75	78	13.
FT	REPEAT	79	82	14.
FT	REPEAT	83	86	15.
FT	REPEAT	87	90	16.
FT	REPEAT	91	94	17.
FT	REPEAT	96	98	18.
SQ	SEQUENCE	104 AA;	9026 MW;	2799D681BFDCC725 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 104;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 69 GGGGS 73

RESULT 5

FUS1_HUMAN

ID FUS1_HUMAN STANDARD; PRT; 110 AA.
 AC O75896;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Fus-1 protein (Fusion 1 protein).
 GN FUS1 OR LGCC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21477323; PubMed=11593436;
 RA Kondo M., Ji L., Kamibayashi C., Tomizawa Y., Randle D., Sekido Y.,
 RA Yokota J., Kashuba V., Zabarovsky E., Kuzmin I., Lerman M., Roth J.,
 RA Minna J.D.;
 RT "Overexpression of candidate tumor suppressor gene FUS1 isolated from
 RT the 3p21.3 homozygous deletion region leads to G1 arrest and growth
 RT inhibition of lung cancer cells.";
 RL Oncogene 20:6258-6262(2001).
 RN [2]
 RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: May function as a tumor suppressor, inhibiting colony
 CC formation, causing G1 arrest and ultimately inducing apoptosis in
 CC homozygous 3p21.3 120-kb region-deficient cells.
 CC -!- TISSUE SPECIFICITY: Strong expression in heart, lung, skeletal
 CC muscle, kidney, and pancreas, followed by brain and liver, lowest
 CC levels in placenta.
 CC -!- SIMILARITY: BELONGS TO THE FUS1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF055479; AAC35497.1; -.
 DR EMBL; BC023976; AAH23976.1; -.
 DR MIM; 607052; -.
 DR GO; GO:0008283; P:cell proliferation; TAS.
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 KW Anti-oncogene.
 SQ SEQUENCE 110 AA; 12074 MW; 9503BD10637C1504 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 110;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 19 GGGGS 23

RESULT 6
 Y0B9_STRCO

ID Y0B9_STRCO STANDARD; PRT; 115 AA.
 AC Q9XAI3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical UPF0133 protein SCO3619.
 GN SCO3619 OR SC66T3.30C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Pelaez A.I., Ribas-Aparicio R.M., Gomez A., Rodicio M.R.;
 RT "Nucleotide sequence analysis and genetic characterization of the recR
 RT gene of Streptomyces."
 RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SIMILARITY: Belongs to the UPF0133 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF151381; AAD34031.2; -.
 DR EMBL; AL939117; CAB45486.1; -.
 DR PIR; T35387; T35387.
 DR HAMAP; MF_00274; -; 1.
 DR InterPro; IPR004401; Cons_hypoth103.
 DR Pfam; PF02575; DUF149; 1.
 DR TIGRFAMs; TIGR00103; TIGR00103; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 115 AA; 11729 MW; 668E1CEDC2E8EB75 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 115;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 104 GGGGS 108

RESULT 7

CRYP_CRYPA

ID CRYP_CRYPA STANDARD; PRT; 118 AA.
AC P52753;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cryparin precursor.
GN CRP.
OS Cryphonectria parasitica (Chesnut blight fungus) (Endothia
OS parasitica).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Diaporthales; Valsaceae;
OC Cryphonectria-Endothia complex; Cryphonectria.
OX NCBI_TaxID=5116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-73.
RC STRAIN=155/2;
RX MEDLINE=94156182; PubMed=8112589;
RA Zhang L., Villalon D., Sun Y., Kazmierczak P., van Alfen N.K.;
RT "Virus-associated down-regulation of the gene encoding cryparin, an
RT abundant cell-surface protein from the chestnut blight fungus,
RT Cryphonectria parasitica.";
RL Gene 139:59-64(1994).
CC -!- FUNCTION: CONTRIBUTES TO SURFACE HYDROPHOBICITY, WHICH IS
CC IMPORTANT FOR PROCESSES SUCH AS ASSOCIATION OF HYPHAE IN
CC REPRODUCTIVE STRUCTURES, DISPERSAL OF AERIAL SPORES AND ADHESION
CC OF PATHOGENS TO HOST STRUCTURES. PRODUCED ABUNDANTLY, EXCEPT IN
CC THE DS-RNA VIRUS-INFECTED STRAINS, WHERE THE EXPRESSION IS MUCH
CC REDUCED.
CC -!- SUBCELLULAR LOCATION: CELL WALL OF AERIAL HYPHAE AND SPORULATION
CC STRUCTURES. ABUNDANTLY SECRETED.
CC -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED ON DAY 2 AND 3 AFTER
CC INOCULATION, A TIME WHEN THE FUNGUS IS IN A RAPID PHASE OF GROWTH.
CC AFTER A STATIONARY PHASE ON DAY 4, THE EXPRESSION DECREASES.
CC -!- SIMILARITY: Belongs to the cerato-ulmin hydrophobin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L09559; AAA19638.1; -.
KW Cell wall; Signal; Repeat.
FT SIGNAL 1 22
FT CHAIN 23 118 CRYPARIN.
FT DOMAIN 23 32 POLY-GLY.

FT	DOMAIN	29	42	7 X 2 AA TANDEM REPEAT OF S-G.
FT	REPEAT	29	30	1.
FT	REPEAT	31	32	2.
FT	REPEAT	33	34	3.
FT	REPEAT	35	36	4.
FT	REPEAT	37	38	5.
FT	REPEAT	39	40	6.
FT	REPEAT	41	42	7.
SQ	SEQUENCE	118 AA; 11387 MW; F7C7CCEEA57D06A5 CRC64;		

Query Match 100.0%; Score 28; DB 1; Length 118;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 25 GGGGS 29

RESULT 8

GRP1_CHERU

ID GRP1_CHERU STANDARD; PRT; 144 AA.
 AC P11898;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Glycine-rich protein HCl.
 OS Chenopodium rubrum (Red goosefoot) (Pigweed).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllales; Amaranthaceae; Chenopodium.
 OX NCBI_TaxID=3560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89240041; PubMed=2717413;
 RA Kaldenhoff R., Richter G.;
 RT "Sequence of cDNA for a novel light-induced glycine-rich protein."
 RL Nucleic Acids Res. 17:2853-2853(1989).
 CC -!- INDUCTION: By light.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X14067; CAA32230.1; -.
 DR PIR; S04069; S04069.
 KW Repeat; Transmembrane.

FT	TRANSMEM	5	25	POTENTIAL.
FT	DOMAIN	37	113	11 X 6 AA TANDEM REPEATS OF G-Y-[NH]-N-G
FT				-G.
FT	REPEAT	37	42	1.
FT	REPEAT	43	48	2.
FT	REPEAT	50	55	3.

FT	REPEAT	56	61	4.
FT	REPEAT	63	68	5.
FT	REPEAT	69	74	6.
FT	REPEAT	76	81	7.
FT	REPEAT	82	87	8.
FT	REPEAT	89	94	9.
FT	REPEAT	102	107	10.
FT	REPEAT	108	113	11.
SQ	SEQUENCE	144 AA; 14137 MW; 5B4D62D4A61621B0 CRC64;		

Query Match 100.0%; Score 28; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 111 GGGGS 115

RESULT 9

GRP9_DAUCA

ID GRP9_DAUCA STANDARD; PRT; 144 AA.
 AC P37703;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Glycine-rich protein DC9.1.
 OS Daucus carota (Carrot).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
 OC Daucus.
 OX NCBI_TaxID=4039;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Aleith F., Richter G.;
 RT "Gene expression during induction of somatic embryogenesis in carrot
 cell suspensions.";
 RL Planta 183:17-24(1990).
 DR PIR; S35716; S35716.
 KW Repeat; Transmembrane.
 FT TRANSMEM 5 25 POTENTIAL.
 FT DOMAIN 37 113 11 X 6 AA TANDEM REPEATS OF G-Y-[NH]-N-G
 FT -G.
 FT REPEAT 37 42 1.
 FT REPEAT 43 48 2.
 FT REPEAT 50 55 3.
 FT REPEAT 56 61 4.
 FT REPEAT 63 68 5.
 FT REPEAT 69 74 6.
 FT REPEAT 76 81 7.
 FT REPEAT 82 87 8.
 FT REPEAT 89 94 9.
 FT REPEAT 102 107 10.
 FT REPEAT 108 113 11.
 SQ SEQUENCE 144 AA; 14111 MW; 5B4D62CFBCA791B0 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
|||||
Db 111 GGGGS 115

RESULT 10

GVA1_STRCO

ID GVA1_STRCO STANDARD; PRT; 144 AA.
AC Q9ZC13;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable gas vesicle structural protein 1 (GVP).
GN GVP1 OR GVPA OR SCO6500 OR SC1E6.09.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabbinoiwitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- FUNCTION: Gas vesicles are small, hollow, gas filled protein
CC structures that are found in several microbial planktonic
CC microorganisms. They allow the positioning of the organism at
CC the favorable depth for growth. GvpA type proteins form the
CC essential core of the structure.
CC -!- SUBCELLULAR LOCATION: Gas vesicle membrane.
CC -!- SIMILARITY: Belongs to the gas vesicle protein type A family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL939128; CAA22037.1; -.
DR PIR; T34730; T34730.
DR HAMAP; MF_00576; -; 1.

DR InterPro; IPR000638; Gas_vesicle.
 DR Pfam; PF00741; Gas_vesicle; 1.
 DR ProDom; PD003598; Gas_vesicle; 1.
 DR PROSITE; PS00234; GAS_VESICLE_A_1; 1.
 DR PROSITE; PS00669; GAS_VESICLE_A_2; FALSE_NEG.
 KW Gas vesicle; Complete proteome.
 SQ SEQUENCE 144 AA; 15315 MW; D1191338B63AFCE6 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 10 GGGGS 14

RESULT 11

PUIB_WHEAT

ID PUIB_WHEAT STANDARD; PRT; 148 AA.
 AC Q10464;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Puroindoline-B precursor.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Capitole; TISSUE=Seed;
 RX MEDLINE=94272013; PubMed=7516201;
 RA Gautier M.-F., Aleman M.-F., Guirao A., Marion D., Joudrier P.;
 RT "Triticum aestivum puroindolines, two basic cystine-rich seed
 RT proteins: cDNA sequence analysis and developmental gene expression."
 RL Plant Mol. Biol. 25:43-57(1994).
 RN [2]
 RP SEQUENCE OF 30-148.
 RA Blochet J.E., Kaboulou A., Compoint J.P., Marion D.;
 RL (In) Bushuk W., Tkachuk R. (eds.);
 RL Gluten proteins, pp.314-325, American Association of Cereal Chemists,
 RL St. Paul (1991).
 CC -!- FUNCTION: Acts as a membranotoxin, probably through its
 CC antibacterial and antifungal activities, contributing to the
 CC defense mechanism of the plant against predators.
 CC -!- PTM: Five disulfide bonds are present.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X69912; CAA49537.1; -.
 DR PIR; S46514; S46514.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR006106; Amylase_inhib.
 DR Pfam; PF00234; tryp_alpha_amyl; 1.
 DR PRINTS; PR00808; AMLASEINHBTR.
 DR SMART; SM00499; AAI; 1.
 KW Plant defense; Membrane; Toxin; Antibiotic; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT PROPEP 20 29
 FT CHAIN 30 148 PUROINDOLINE-B.
 FT DOMAIN 68 73 TRP-RICH.
 SQ SEQUENCE 148 AA; 16792 MW; 327904B4EBEC2C16 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 148;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGS 5
 |||||
 Db 33 GGGGS 37

RESULT 12

GRP_DAUCA

ID GRP_DAUCA STANDARD; PRT; 157 AA.
 AC Q03878;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glycine-rich RNA-binding protein.
 OS Daucus carota (Carrot).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
 OC Daucus.
 OX NCBI_TaxID=4039;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Queen Anne's Lace;
 RA Sturm A.;
 RT "A wound-inducible glycine-rich protein from Daucus carota with
 RT homology to single-stranded nucleic acid binding proteins.";
 RL Plant Physiol. 99:1689-1692(1992).
 CC -!- FUNCTION: May play a role in the biosynthesis and processing of
 CC heterogeneous nuclear RNA and in the maturation of specific mRNAs
 CC in response to wounding.
 CC -!- INDUCTION: In response to stress by wounding.
 CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; X58146; CAA41152.1; -.
DR PIR; S14857; S14857.
DR HSSP; P09651; 1HA1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW RNA-binding.
FT DOMAIN      6      84      RNA-BINDING (RRM).
FT DOMAIN      86     154      GLY-RICH.
SQ SEQUENCE    157 AA;  15718 MW;  73FBD644F51CB633 CRC64;

```

```

Query Match          100.0%; Score 28; DB 1; Length 157;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches      5; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 GGGGS 5
        |||||
Db      151 GGGGS 155

```

RESULT 13
CGC8_MOUSE

```

ID CGC8_MOUSE      STANDARD;      PRT;  163 AA.
AC Q9D187;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical protein CGI-128 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;

```

RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -!- SIMILARITY: Belongs to the UPF0195 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AK003830; BAB23024.1; -.
 DR MGD; MGI:1915773; 1110019N10Rik.
 DR InterPro; IPR002744; DUF59.
 DR Pfam; PF01883; DUF59; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 163 AA; 17667 MW; D3171D52CF3AD02F CRC64;

Query Match 100.0%; Score 28; DB 1; Length 163;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
 |||||
 Db 3 GGGGS 7

RESULT 14

SSB_BRAJA

ID SSB_BRAJA STANDARD; PRT; 164 AA.
 AC Q89L50;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Single-strand binding protein (SSB) (Helix-destabilizing protein).
 GN SSB OR BLL4698.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197(2002).
 CC -!- FUNCTION: This protein is essential for replication of the
 CC chromosome. It is also involved in DNA recombination and repair
 CC (By similarity).
 CC -!- SIMILARITY: Contains 1 SSB domain.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AP005952; BAC49963.1; -.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR InterPro; IPR000424; SSB_protein.
DR Pfam; PF00436; SSB; 1.
DR PROSITE; PS50935; SSB; 1.
KW DNA-binding; DNA repair; DNA replication; Complete proteome.
FT DOMAIN 5 111 SSB.
FT DOMAIN 116 123 POLY-GLY.
SQ SEQUENCE 164 AA; 17522 MW; 277330E21E8418BA CRC64;

Query Match 100.0%; Score 28; DB 1; Length 164;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5
| | | |
Db 120 GGGGS 124

RESULT 15

GRP1_ORYSA

ID GRP1_ORYSA STANDARD; PRT; 165 AA.
AC P25074;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Glycine-rich cell wall structural protein 1 precursor.
GN GRP-1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR36;
RX MEDLINE=91370862; PubMed=1716496;
RA Lei M., Wu R.;
RT "A novel glycine-rich cell wall protein gene in rice."
RL Plant Mol. Biol. 16:187-198(1991).
CC -!- FUNCTION: Responsible for plasticity of the cell wall (Potential).
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; X53596; CAA37665.1; -.

DR PIR; S13385; KNRZG1.

DR Gramene; P25074; -.

KW Cell wall; Structural protein; Repeat; Signal.

FT SIGNAL 1 23 POTENTIAL.

FT CHAIN 24 165 GLYCINE-RICH CELL WALL STRUCTURAL

FT PROTEIN 1.

FT DOMAIN 31 159 GLY-RICH.

FT REPEAT 56 62 R2 (TYR-RICH).

FT REPEAT 93 99 R2 (TYR-RICH).

FT REPEAT 132 138 R2 (TYR-RICH).

SQ SEQUENCE 165 AA; 13536 MW; E36CE31C3650AC9A CRC64;

Query Match 100.0%; Score 28; DB 1; Length 165;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGGS 5

||||

Db 121 GGGGS 125

Search completed: March 5, 2004, 16:23:44

Job time : 1.67901 secs